

QY 2561 IleIysSerValTyrIleuValLeuGluAspSerGluThrProIleProThrThrIle 2580
 DB 8021 ATCAAGTCGCTGGAGAGACTCTCTGAGACTCGAAGAACCACTATCCCAACCACT 8080
 QY 2581 MetAlaIysAsnGluValPheCysValAspProThrIleGlyIleValAlaAlaArg 2600
 DB 8081 ATGGCCAAATATGAGGTCTCTGCGTGGACCCCAACAGGGGGGAGAAAGAGCTGCG 8140
 QY 2601 LeuIleValTyrProAspLeuGlyValArgValCysGluIleMetAlaLeuTyrAspIle 2620
 DB 8141 CTTATCGTTTACCTGACTCGGCGCTCAGGCTCTGGAGAGAGATGCGCTTTATGACAT 8200
 QY 2621 ThrGluIleLeuProGluAlaValMetGlyValAspSerGlyPheGluTyrSerProAla 2640
 DB 8201 ACACAAATCTCTCAGCGCGTGAATGGGGCTCTTATGGAATTCAGGTATTCCTCCGCT 8260
 QY 2641 GluArgValGluPheIleLeuIleValAlaIleGluIleValAspProMetGlyPheSer 2660
 DB 8261 CAGCGGTAGAGTTCTCTGAAGCATGCGCGGAGAAAGAGAGCCCTATGGGTATTTTCG 8320
 QY 2661 TyrAspThrArgCysPheAspSerThrValThrGluArgAspIleArgThrGluGluSer 2680
 DB 8321 TATCATACCCGATGCTTTGACTCAACCGTCACTGAGAGAGACATCAGGACTCAGGAGTCC 8380
 QY 2681 IleTyrArgAlaCysSerLeuProGluGluAlaIleThrAlaIleHisSerLeuThrGlu 2700
 DB 8381 ATATATGGGCTGCTCTCTCCCGAGGAGGCCACACTGCCATACACTCGCTAACTGAG 8440
 QY 2701 ArgLeuTyrValGlyIleProMetPheAsnSerIleGlyGluThrCysGlyTyrArgArg 2720
 DB 8441 AGACTTAGTGGAGGGGCTATGTTCAACGACAGGGGCTCAACCTCGGGTACAGGCT 8500
 QY 2721 CysArgAlaSerGlyValLeuThrThrSerMetGlyAsnThrIleThrCysArgValIys 2740
 DB 8501 TGGCGGCCAGCGGGGTCTACCACTAGCATGAGAGAGAGACATCATCATCATCATCAT 8560
 QY 2741 AlaLeuAlaCysIleValAlaAlaGlyIleIleAlaProThrMetLeuValCysGlyAsp 2760
 DB 8561 GCCTTAGCGCTGTGAAGCTGCAGGGATATATCGGCCCACTATCGTGTATGGCGGAT 8620
 QY 2761 AspLeuValIleSerGluSerGluGlyThrGluGluAspGluArgAsnLeuArgAla 2780
 DB 8621 GACTTGGTGTCTATCTCAGAAAGTCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 8680
 QY 2781 PheThrGluAlaMetThrArgTyrSerAlaProProGluAspProArgProGluTyr 2800
 DB 8681 TTCAGGAGGCTATGACGAGGTATTCCTGCTCTCTGCTGACCCCTCCAGACCGGATAT 8740
 QY 2801 AspLeuGluIleThrSerCysSerSerAsnValSerValAlaLeuGlyProGluGly 2820
 DB 8741 GATCTGAGCTGATACATCTTGTCTCTCAAAATGTGTCTGGCGCTGGCGCCCAAGGC 8800
 QY 2821 ArgArgGlyTyrIleuThrArgAspProThrProIleAlaArgAlaAlaIleProGlu 2840
 DB 8801 CGCGCAGATATCTACCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8860
 QY 2841 ThrValArgHisSerProValAsnSerIleLeuGluValAsnIleIleGlnTyrAlaProThr 2860
 DB 8861 ACAGTTAGACACTCCCTCTCAATTCATGCTGGAGAGAGAGAGAGAGAGAGAGAGAG 8920
 QY 2861 IleThrAlaArgMetValLeuMetThrIlePhePheSerIleLeuMetAlaGlnAspThr 2880
 DB 8921 ATATGGCTCGCATGCTCTGATGACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8980
 QY 2881 LeuAspGlnAsnLeuAsnPheGluMetTyrGlyAlaValTyrSerValSerProLeuAsp 2900
 DB 8981 CTGACAG 9040
 QY 2901 LeuProAlaIleIleGluArgLeuHisGlyIleuAspAlaPheSerLeuHisThrThr 2920
 DB 9041 CTCGAGCTATATTAAGAGGTATCATGAGGCTTTCTCTCTCTCTCTCTCTCTCTCTCT 9100
 QY 2921 ProHisGluLeuThrArgValAlaSerAlaLeuArgIysLeuGlyAlaProLeuArg 2940

DB 9101 CCCACGAACCTGACCGGTGGCTTTCAGCCCTCAGAAACTTGGGGCCACCCCTCAGA 9160
 QY 2941 AlATrPlySerArgAlaArgAlaValArgAlaSerIleuLeuSerArgGlyGlyArgAla 2960
 DB 9161 GCCTGGAAGAGCGGGCAGTGCAGTCAAGGCGCTCTCATCTCCGCTGGGGGAGAGCG 9220
 QY 2961 AlaValCysGlyValTyrTyrIleuPheAsnTrpAlaValIleThrIleuLeuLeuThrPro 2980
 DB 9221 GCGTITGGCGTGCATATCTCTTCAATGGGCGGTGAAGACCAAGCTCAAACTCACTCCA 9280
 QY 2981 LeuProGluAlaArgLeuLeuAspLeuSerSerTrpPheThrValGlyAlaGlyGlyGly 3000
 DB 9281 TTCGGGAGAGCGGCTCTCTGATTTATCCAGCTGGTTCACCGTGGGCGCGGGGCG 9340
 QY 3001 AspIleTyrHisSerValSerArgAlaArgProArgLeuLeuPheGlyLeuLeuLeu 3020
 DB 9341 GACATTTATCAGAGGCTGTCGCTGCGGACCGCTTATGCTCTTTGGCTTACTCTTA 9400
 QY 3021 LeuPheValGlyValGlyIleuPheLeuLeuProAlaArg 3033
 DB 9401 CTTTTTGTAGGGTAGGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 9439
 RESULT 4
 AAQ38218
 ID AAQ38218 standard; cDNA to mRNA; 9589 BP.
 XX AC AAQ38218;
 XX DT 25-MAR-2003 (updated)
 XX DT 01-JUL-1993 (first entry)
 XX DE NANBH virus strain HC-J6 genome cDNA sequence.
 XX KW Non A non B hepatitis virus; amplification; HC-J1; HC-J8; plasma; ss.
 XX OS Non A, non B hepatitis virus strain HC-J6.
 XX FH Key Location/Qualifiers
 FT CDS 341..9442
 FT /*tag = a
 XX PN EP532167-A2.
 XX PD 17-MAR-1993.
 XX PF 30-JUL-1992; 92EP-0306952.
 XX PR 09-AUG-1991; 91JP-0287402.
 XX PR 05-DEC-1991; 91JP-0360441.
 XX PA (IMMO) IMMUNO JAPAN INC.
 XX PI Nakamura, T., Okamoto, H.;
 XX PI WPI, 1993-087166/11.
 XX DR N-PSDB; AAR33538.
 XX PT Polynucleotide(s), polypeptide(s) and antibodies of NANBH virus -
 PT useful for detecting NANBH, as a vaccine and for screening blood
 PT samples
 XX PS Claim 2; Page 27-32; 93pp; English.
 XX CC RNA was isolated from the plasma of human patients positive for
 CC NANBH virus (strain HC-J6) and was subjected to reverse transcription
 CC to produce cDNA. The resulting cDNA was amplified by PCR, and
 CC nucleic acid sequences determined by analysis of both clones from the
 CC cDNA library and clones obd. by PCR amplification (36 clones in total).
 CC The NANBH HC-J6 genome was found to contain an open reading frame
 CC encoding a polypeptide precursor of 3033 amino acid residues.
 CC See also AAQ38172-221.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 9589 BP: 1968 A: 2820 C: 2635 G: 2166 T: 0 other;

Alignment Scores:

Pred. No.: 0 Length: 9589
 Score: 15824.00 Matches: 2966
 Percent Similarity: 98.788 Conserved: 30
 Best Local Similarity: 97.798 Mismatches: 37
 Query Match: 98.168 Indels: 0
 DB: 14 Gaps: 0

US-09-980-559-2 (1-3033) x AAQ38218 (1-9589)

Qy 1 MetSerThrAsnProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
 Db 341 ATGAGCACAAATCTTAAACCTCAAGAGAAAACCAAGAAACACCAACCGTGGCCACA 400
 Qy 21 AspValLysPheProGlyGlyGlnLysValGlyValLysValLysValLysVal 40
 Db 401 GACGTTAAGTTTCGGGGGGGGCCAGATCGTTGGCGGAGTATCTTTCGGGGCAGG 460
 Qy 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 60
 Db 461 GGGCCCCAGGTGGGTGGTGGCGCCGACAGAGAGAGACTTCGGAGGGGTCCCGAGCCACGTGGA 520
 Qy 61 ArgArgGlnProLysProLysAspArgArgSerThrGlyLysSerThrGlyLysProGly 80
 Db 521 AGCGCGCAGCCCTCCCTAAGGATCGCGCGCTCCACTGGCCAAATCTTCGGGGAACACAGGA 580
 Qy 81 TyrProTrpProLeuTyrGlyAsnGluGlyLeuGlyTrpAlaGlyTrpLeuLeuSerPro 100
 Db 581 TACCCCTGGCCCTATACGGGATCAGGGACTCGGCTGGCGGAGATGGCTCTGTCCCC 640
 Qy 101 ArgGlySerArgProSerTrpGlyProAsnAspProHishArgSerArgAsnValGly 120
 Db 641 CGAGTTCGGTCCCTCTTCGGGGCCCCATGACCCCCGGCATAGGTCCCGCAACGTGGGT 700
 Qy 121 LysValLysAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProValVal 140
 Db 701 AAGGTCATCGATACCTTAACGTGGCGGCTTTGGCGACCTCATGGGGTACATCTCTCGTA 760
 Qy 141 GlyValProLeuGlyGlyValAlaArgAlaLeuAlaHisGlyValArgValLeuGluAsp 160
 Db 761 GGGCGCCCGCTCGCGGGCGTGGCCAGAGCTCTGGCGCATGGCTGAGAGTCTGGAGGAC 820
 Qy 161 GlyValAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeuAla 180
 Db 821 GGGGTTAATTTTGCACAGAGGAACTTACCCGGTTCCTCTTCTATCTTCTTCTGCTGCC 880
 Qy 181 LeuLeuSerCysIleThrThrProValSerAlaAlaGluValLysAsnIleSerThrGly 200
 Db 881 CTGCTGTCTGTCATCACCACCCCGGTCTCGCTGCCGAGTGAAGAAACATCAGTACCAGG 940
 Qy 201 TyrMetValThrAsnAspCysThrAsnAspSerIleThrTrpGlnLeuGlnAlaAlaVal 220
 Db 941 TACATGGTGCACCAACGATGCACCAATGATAGCATTACCTGGCACTCCAGGCTGCTGTC 1000
 Qy 221 LeuHisValProGlyCysValProCysGluLysValGlyAsnAlaSerGlnCysTrpIle 240
 Db 1001 CTCACGTCCTCCGGTGGGTGGTGGTGGAGAAAGTGGGGAAATACATCTCGGTCTGGA 1060
 Qy 241 ProValSerProAsnValAlaValGlnArgProGlyAlaLeuThrGlnGlyLeuArgThr 260
 Db 1061 CCGGTCTCACCGAAATGTGGCGGTGCAGCAGCCCGCGCCCTCAGCAGAGGGCTTACGAG 1120
 Qy 261 HisIleAspMetValValMetSerAlaThrLeuCysSerAlaLeuTyrValGlyAspLeu 280
 Db 1121 CACATTCACATGATGTGATGTCCGCCACGCTCTGCTCTTTCAGTGGGGGACCTC 1180
 Qy 281 CysGlyGlyValMetLeuAlaGlnMetPheIleValSerProGlnHisIleTrpPhe 300
 Db 1181 TCGGGTGGGTGATGCTTCGAGCCCGAGATTTTATATGTCTCGCCACAGCACCACTGGTTT 1240

Qy 301 ValGlnAspCysAsnCysSerIleTyrProGlyThrIleThrGlyHisArgMetAlaTrp 320
 Db 1241 GTGCAAGACTGCATTTGCTCATCTACCTTGGTACCATCCTGGACACCGCATGGGTGG 1300
 Qy 321 AspMetMetMetAsnTrpSerProThrAlaThrMetIleLeuAlaTyrAlaMetArgVal 340
 Db 1301 GACATGATGATACTGTGTCGCCCGCTACCATGCTCTGGCGGTACCGCATGGGTGG 1360
 Qy 341 ProGluValIleIleLeuPheIleSerGlyAlaHisIleTrpGlyValMetPheGlyLeuAla 360
 Db 1361 CCCGAGGTCTCATPAGACATCATTTGGCGGGGCTCATTTGGGGCGTCATGTCGGTTAGCC 1420
 Qy 361 TyrPheSerMetGlnGlyAlaTrpAlaLysValValValIleLeuLeuLeuAlaGly 380
 Db 1421 TACTTCTCTATGAGGAGCGTGGCCAAAGTGGTGTCTTCTTCTTCTTCTTCTTCTTCTTCT 1480
 Qy 381 ValAspAlaArgThrHisThrValGlyGlySerAlaAlaGlnThrThrGlyArgLeuThr 400
 Db 1481 GTGAGCGGCAACCCATACCTTGGGGGTTTTCACGGCGCATACGCCGAGACCTTACC 1540
 Qy 401 SerLeuPheAspMetGlyProArgGlnLysIleGlnLeuValAsnThrAsnGlySerTrp 420
 Db 1541 GGCATGTTCTCCCTTGGTGGCCAGGAGAAATCCAGCTCATCAACCAATGGCAGTTGG 1600
 Qy 421 HisIleAsnArgThrAlaLeuAsnCysAsnAspSerLeuHisThrGlyPheIleAlaSer 440
 Db 1601 CACATCAACCGCACCGCCCTGAATGCAATGACTCTTTGCACACCGGCTCTCTCCGCTCA 1660
 Qy 441 LeuPheTyrThrHisSerPheAsnSerSerGlyCysProGluAspMetSerAlaCysArg 460
 Db 1661 CTGTTCTACACCCACAGCTTCACTGTTCAGGATGTTCGGAACGATGTCCGCTTCGCG 1720
 Qy 461 SerIleGluAlaPheArgValGlyTrpGlyAlaLeuGlnTyrGluAspAsnValThrAsn 480
 Db 1721 AGTATCGAGGCTTTTGGGTGGGATGGGGCGCTTACCAATATGAGGACAAATGTCCCAAT 1780
 Qy 481 ProGluAspMetArgProTyrCysTrpHisTyrProProArgGlnCysGlyValValSer 500
 Db 1781 CCGAGGATATGAGACCGTATTCTGCGCACTACCCACCAACAGACAGTGTGTGTAGTCTCC 1840
 Qy 501 AlaLysThrValCysGlyProValTyrCysPheThrProSerProValValValGlyThr 520
 Db 1841 GCGAGCTCTGTGTGGCCAGTGTACTGTTCACCCCGCCAGCCAGTGTGTGGTAGG 1900
 Qy 521 ThrAspArgLeuGlyAlaProThrTyrThrTrpGlyGluAsnGluThrAspValPheLeu 540
 Db 1901 ACCGATAGACTGTGAGCGCCCTTACACGTGGGGGAGAAATGAGACAGATGTCTTCTTA 1960
 Qy 541 LeuAsnSerThrArgProProLeuGlySerTrpPheGlyCysThrTrpMetAsnSerSer 560
 Db 1961 TTGAACAGCACTCGACCCAGCGAGGGGTTCATGGTTGGCGTGCACGTGTGATCACTCCACT 2020
 Qy 561 GlyTyrThrLysThrCysGlyAlaProProCysArgThrArgAlaAspPheAsnAlaSer 580
 Db 2021 GGCTACACCAAGACTTGGCGCGCACCCCTCGCGCATTTAGAGCTGACTTCAATGCCAGC 2080
 Qy 581 ThrAspLeuLeuCysProThrAspCysPheArgLysHisProAspThrThrTyrLeuLys 600
 Db 2081 ATGAGCTGTGTGGCCCGCCAGCTGTGTTAGAGAGCATCTGTATACCACTACATCAAA 2140
 Qy 601 CysGlySerGlyProTrpLeuThrProArgCysLeuIleAspTyrProTyrArgLeuTrp 620
 Db 2141 TGTGGCTCTGGCGCTGGCTCGCCAGCGGAGTGTGATCGATCGATCCCTTACAGGCTCTGG 2200
 Qy 621 HisTyrProCysThrValAsnTyrThrIlePheLysIleArgMetTyrValGlyVal 640
 Db 2201 CATTACCCCTGCACAGTTAATACCATCTTCAAAATAAGGATGTATGTGGGGGGGCTC 2260
 Qy 641 GluHisArgLeuThrAlaAlaCysAsnPheThrArgGlyAspAspArgCysAsnLeuGluAsp 660
 Db 2261 GAGCAGAGGCTCAGCGCTGGTGCAATTTCACTGTGGGGATCGTGTGCACTTGGAGGAC 2320
 Qy 661 ArgAspArgSerGlnLeuSerProLeuLeuHisSerThrThrGluTrpAlaIleLeuPro 680

3401	Db	GCCTATACCTCCAAAGGGGTGGAGTCTTCTGCGCCCCCATCACTGCTTATGCGCCACGACGACA	3460
1041	Qy	ArgGlyLeuLeuGlyThrIleValValSerMetThrGlyArgAspIysThrGluGlnAla	1060
3461	Db	CCGGCCCTTTTGGGCACCAATAGTGGGTGACATGACGCGGGCCGACACAGACAGCAACAGCGCC	3520
1061	Qy	GlyGluIleGlnValLeuSerThrValThrGlnSerPheLeuGlyThrSerIleSerGly	1080
3521	Db	GGCGAGATTCAAGTCTCTGTCACCGGTCACTCAGTCTCTCTCGGACCAACCAATCTCTCGGG	3580
1081	Qy	ValLeuTrpThrValTyriHisGlyAlaGlyAsnIysValThrLeuAlaGlySerArgGlyPro	1100
3581	Db	GTCTTATGGACTGTCTACCAATGGAGCTGGCAACAGACTCTAGCCGGCTCACGGGGTCGG	3640
1101	Qy	ValThrGlnMetTyrSerSerAlaGlyAspLeuValGlyTyrProSerProProGly	1120
3641	Db	GTACACAGATGTACTCCAGTGTCTAGCGGGGACTTAGTGGGGGTGGGCCAGCCCCCCCCGG	3700
1121	Qy	ThrIysSerLeuGluProCysThrCysGlyAlaValAspLeuTyrLeuValThrArgAsn	1140
3701	Db	ACCAAAATCTTTGGAGCGGTGCACGTGGAGCGGTGCACCTATACCTGGTCTACGCGAATC	3760
1141	Qy	AlaAspValIleProAlaArgArgGlyAspIysArgGlyAlaLeuLeuSerProArg	1160
3761	Db	GCTGATGTATCTCCGGCTCCGAAGACCGGGGACAAGCGAGAGCGCTACTCTCTCCCGAGA	3820
1161	Qy	ProLeuSerThrLeuLeuIysGlySerSerGlyValProValLeuCysProArgGlyHisAla	1180
3821	Db	CCTCTTCCACCTTGAAGGGTCTCTCGGGGGCCCGGTGCTCTGCGCCACGAGGCCACCGCT	3880
1181	Qy	ValGlyValPheArgAlaAlaValCysSerThrGlyValAlaIysSerIleAspPheIle	1200
3881	Db	GTCCGGGTCTTCCGGCGACCGGTGTCTCCCGGGGGCGTGGCCAAAGTCCATAGATTATTATC	3940
1201	Qy	ProValGluThrLeuAspIleValThrArgSerProThrPheSerAspAsnSerThrPro	1220
3941	Db	CCGGTTTGAGACACTTGACATCGTCACCTCGGTCCGCCACCTTTAGTGACAACAGCACCA	4000
1221	Qy	ProAlaValProGlnThrTyriGlnValGlyTyrLeuHisAlaProThrGlySerGlyIys	1240
4001	Db	CCTGCTGTGCCCCAAACTTATCAGGTGGGTACTTACATGCCCCCGACTGCTAGTGTGGAAG	4060
1241	Qy	SerThrIysValProValAlaTyriAlaAlaGlnGlyTyrIvalLeuValLeuAsnPro	1260
4061	Db	AGCACCAAAAGTCCCTGTCCGGTATGCCGCTCAGGGGTACAAAGTGTCTAGTGTCTAATGCC	4120
1261	Qy	SerValAlaAlaThrLeuGlyPheGlyValaTyriLeuSerIysValHisGlyIleAsnPro	1280
4121	Db	TCGGTGGCTGCCACCCTCGGGGTTTGGGGCGTACTTGTTCGAAGCACATGGCATCAATGCC	4180
1281	Qy	AsnIleArgThrGlyValArgThrValIleThrThrGlyAlaProIleThrTyriSerThrTyri	1300
4181	Db	AACATTAGACATGGGGTTCAGACATGTGACGACCGGGGCGCCCATCACGTACTCCACATAT	4240
1301	Qy	GlyIysPheLeuAlaAspGlyGlyCysAlaGlyGlyValaTyriAspIleIleIleCysAsp	1320
4241	Db	GGCAAAATCTCTCCCGATGGGGGTGCGCGGGCGGCGCTATGACATCATCATATGCGAT	4300
1321	Qy	GluCysHisAlaValAspSerThrThrIleLeuGlyIleGlyThrValLeuAspGlnAla	1340
4301	Db	GAATGCCATGCCGTGGACTCTTACCAACCAATCTCTGGCATCGGAACAGTCTCTCGATCMAGCA	4360
1341	Qy	GluThrAlaGlyValArgLeuThrValLeuAlaThrAlaThrProProGlySerValThr	1360
4361	Db	GAGACGCGGGGTCAAGGCTAACTGTACTGGCTACGGCTACGCCGCCCGGGTCAGTAGACA	4420
1361	Qy	ThrProHisProAsnIleGluValAlaLeuGlyGlnGlnGlyIleProPheTyri	1380
4421	Db	ACCCCCACCCCAKATAGAGAGGTGGCCCTCGGGCAGAGGGGTGAGATGCCCTTCTAT	4480
1381	Qy	GlyArgAlaIleProLeuSerTyriIleLysGlyGlyIysHisLeuIlePheCysHisSer	1400
4481	Db	GGGAGGGGCAATTCCTCTGTCTATCATCAAGGAGGAGAACACTTATCTCTCTGCCACTCA	4540

Qy 1401 LysLeuValGlyCysAspGluLeuAlaAlaAlaLeuArgGlyMetGlyLeuAsnSerValAla 1420
 Db 4541 AAGAAAGGTGACAGCTCGCGCGCTTCGGGGTATGGCTTGAGCCAGTGGCA 4600
 Qy 1421 TyrTyrArgGlyLeuAspValSerValIleProThrGlnGlyAspValValValAla 1440
 Db 4601 TACTACAGAGGCTCGAGCTCTCGTAATACCAACTCAGGGAGCCTAGTGTGTGCGCC 4620
 Qy 1441 ThrAspAlaLeuMetThrGlyTyrThrGlyAspPheAspSerValIleAspCysAsnVal 1460
 Db 4661 ACCGACGCTCATGACGGGTCTTACTGGAGCTTTGACTCTCGTGTGCTGCAACCTA 4720
 Qy 1461 AlaValThrGlnValValAspPheSerLeuAspProThrPheThrIleThrThrGlnIle 1480
 Db 4721 GCGGTCACTCAAGTGTGAGCTTCAGCTTGACCCCACTTACCAATACCAAGCACT 4780
 Qy 1481 ValProGlnAspAlaValSerArgSerGlnArgGlyArgThrGlyArgGlyArgLeu 1500
 Db 4781 GTCCCTCAAGAGCTGTCTCAGTAGCAGCGCGCGCGCGCACGCGGAGGAGAGCTG 4840
 Qy 1501 GlyIleTyrArgTyrValSerThrGlyLeuArgAlaSerGlyMetPheAspSerValVal 1520
 Db 4841 GGTATTTATAGCTATGTTTCCACTGTGTGAGCGAGCTTCAGGAATGTTTGCAGTGTAGT 4900
 Qy 1521 LeuCysGluCysTyrAspAlaGlyValAlaIleTyrGluLeuThrProSerGluThrThr 1540
 Db 4901 CTCTGCACTGTCTGAGTGCAGAGGGCGCATGTGTATGTGCTCACACGCGAGACCACT 4960
 Qy 1541 ValArgLeuArgAlaTyrPheAsnThrProGlyLeuProValCysGlnAspHisLeuGlu 1560
 Db 4961 GTCAAGCTCAGAGCATATTTCAACACACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5020
 Qy 1561 PheTyrGluAlaValPheThrGlyLeuThrHisIleAspAlaHisPheLeuSerGlnThr 1580
 Db 5021 TTTTGGAGCGAGTTCACCGGCTCACACATAGATGCCCACTTCCTTTCCCAACA 5080
 Qy 1581 LysGlnSerGlyGluAsnPheAlaTyrLeuThrAlaTyrGlnAlaThrValCysAlaArg 1600
 Db 5081 AAGCAATCGGGGAAATTTCCGACTATTACAGCTTACAGGCTACAGCTGTGCGCTAGG 5140
 Qy 1601 AlaValAspProProSerTyrAspValMetTyrIleGlyLeuThrArgLeuLysPro 1620
 Db 5141 GCCAAGCCCCCCCCCTGTGAGCTGTATGGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5200
 Qy 1621 ThrLeuValGlyProThrProLeuLeuTyrArgLeuGlySerValThrAsnGluValThr 1640
 Db 5201 ACACCTGTGGGCCCCACACCTCTCTGTACCGCTTGGGCTCTGTGTACCAACGAGGTCACT 5260
 Qy 1641 LeuThrHisProValThrLysTyrIleAlaThrCysMetGlnAlaAspLeuGluValMet 1660
 Db 5261 CTCACGATCTCTGTACGGAATACATCGCCACCTGTGATGAGCGGAGCTGTGAGGTCTATG 5320
 Qy 1661 ThrSerThrTyrValLeuAlaGlyGlyValLeuAlaAlaValAlaIleTyrCysLeuAla 1680
 Db 5321 ACCAGACGTGTGGTCTTAGCTGGGGGGGTCTTGGGCGCTGTGCGCTGTGCTGTGCTGTG 5380
 Qy 1681 ThrGlyCysValCysIleIleGlyArgLeuHisIleAsnGlnArgAlaValValAlaPro 1700
 Db 5381 ACCGGGTGTGTGTGATCATCGGCGCTTGCAGCGTTAAACAGCGAGCGCTGTGTGTGTG 5440
 Qy 1701 AspLysGluValLeuTyrGluAlaPheAspGluMetGluCysAlaSerArgAlaAla 1720
 Db 5441 GACAAGGAGTCTCTATGAGGCTTTGATGAGATGAGGGAATGTGCTCTAGAGCGGCT 5500
 Qy 1721 LeuIleGluGluGlyGluArgIleAlaGluMetLeuLysSerIleGlnIleLeu 1740
 Db 5501 CTATTTGAAGAGGGGCGAGATAGCGAGATGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 5560
 Qy 1741 GlnGlnAlaSerLysGlnAlaGlnAspIleGlnProThrValGlnAlaSerTyrProLys 1760
 Db 5561 CAGCAAGCTTCCAAACAGCTCAAGACATACACCGCTGTGAGGCTTCTTGGGCCCAAG 5620

Qy 1761 ValGluGlnPheTyrAlaLysHisMetTyrAsnPheIleSerGlyIleGlnTyrLeuAla 1780
 Db 5621 GTAGAGCATTTCTGGGCGCAAAACACATGTGGAACTTTCATCAGCGGATTTCAATACCTCGCA 5680
 Qy 1781 GlyLeuSerThrLeuProGlyAsnProAlaValAlaLeuSerMetMetAlaPheSerAlaAla 1800
 Db 5681 GGCATATCAACACATCGCCAGGAAACCTGCTGTGTAGCTTCCATCATGATGGCATTTCAAGTGGCGCC 5740
 Qy 1801 LeuThrSerProLeuSerThrThrThrIleLeuLeuAsnIleLeuGlyGlyTyrLeu 1820
 Db 5741 CTCACAGTCCGTGTGTCACTAGCACCATATCTCTTCTCAACATTTTGGGGGGCTGGCTA 5800
 Qy 1821 AlaSerGlnIleAlaProProAlaGlyAlaThrGlyPheValValSerGlyLeuValGly 1840
 Db 5801 GCATCCCAATTTGGGCTCTCCCGGGGGCTACCGGCTTCTGTGTGTGTGTGTGTGTGTGTGT 5860
 Qy 1841 AlaValGlySerIleGlyLeuGlyValValLeuValAspIleLeuAlaGlyTyrGly 1860
 Db 5861 GCTGCGCTAGCGACATAGGCTTGGGTAGGTGTGGTGTGGACATCTCTGGCGAGGTATGGT 5920
 Qy 1861 AlaGlyIleSerGlyAlaLeuValAlaPheLysIleMetSerGlyGlyLysProSerMet 1880
 Db 5921 GCGGCAATTTGCGGGGCTCTGTGTGSCATTTCAAGATCATGTCTGTGGAGAGCCCTCCATG 5980
 Qy 1881 GluAspValValAsnLeuLeuProGlyIleLeuSerProGlyAlaLeuValValGlyVal 1900
 Db 5981 GAGATGTGTGTCAACCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6040
 Qy 1901 IleCysAlaAlaIleLeuArgArgHisValGlyProGlyGlyGlyAlaValGlnTyrMet 1920
 Db 6041 ATCTGCGCGCCCATCTCTGCGCGCACACGTGGGAGCAGGGGAGGCGCTGTCCAAATGGATG 6100
 Qy 1921 AsnArgLeuIleAlaPheAlaSerArgGlyAsnHisValAlaProThrHisTyrValThr 1940
 Db 6101 AATAGGCTCATTTGCTTGTCTTCCAGAGGAAACACAGCTCGGCCCCACCCACTAGCTGACG 6160
 Qy 1941 GluSerAspAlaSerGlnArgValThrGlnLeuLeuGlySerLeuThrIleThrSerLeu 1960
 Db 6161 CAGTCCGATGGTGTGCGAGCGGTGTGACCACTTACTTGGCTCCCTTACCATAACCCAGCTG 6220
 Qy 1961 LeuArgArgLeuHisMetTyrIleThrGluAspCysProIleProCysGlyGlySerTyr 1980
 Db 6221 CTAGAGAGCTTCAACAACTGTATCTGAGACTGCCCATCTCCCATGAGCGGCTCTGTG 6280
 Qy 1981 LeuArgAspValTyrAspTyrValCysThrIleLeuThrAspPheLysAsnTyrLeuThr 2000
 Db 6281 CTCCGCGATGTGTGGATTTGGGTTTGACCATCTTAAACAGACTTTTAAACAACTGGCTGACC 6340
 Qy 2001 SerLysLeuPheProLysMetProGlyLeuProPheValSerCysGlnLysGlyTyrLys 2020
 Db 6341 TCCAAATTTTCCCAAGATGCTGTGCTCTCCCTTTATCTCTTGTCAAAAGGGGTACAG 6400
 Qy 2021 GlyValTyrAlaGlyThrGlyIleMetThrThrArgCysProCysGlyValAsnIleSer 2040
 Db 6401 GCGGTGTGGGCTGGCATGTGTATCATGACACACGCTGTCTTGTGGCGGCCCAATATCTCT 6460
 Qy 2041 GlyAsnValArgLeuGlySerMetArgIleThrGlyProLysThrCysMetAsnIleTyr 2060
 Db 6461 GGCATGTCCGCTGTGGCTCCATGAGATTAACGGGGCCCAAAACCTGTGATGATATCTGG 6520
 Qy 2061 GlnGlyThrPheProIleAsnCysTyrThrGluGlyGlnCysValProLysProAlaPro 2080
 Db 6521 CAGGGAGCTTTTCCATCAATTTTACACGAGGGGCGAGTGTGCTGCGCAAAACCCGACCA 6580
 Qy 2081 AsnPheLysValAlaIleTyrArgValAlaAlaSerGluTyrAlaGlnValThrGlnHis 2100
 Db 6581 AACTTTAGATTCGCCATCTGGAGGGTGGCGGCTCAGAGTACCGGAGGTGACCGACAC 6640
 Qy 2101 GlySerTyrHisTyrIleThrGlyLeuThrThrAspAsnLeuLysValProCysGlnLeu 2120
 Db 6641 GGGTCATACCACTACATAACAGACTTACCACTGTAACTTGAAGTTCCTTGTGCAACTA 6700
 Qy 2121 ProSerProGluPhePheSerTyrValAspGlyValGlnIleHisArgPheAlaProThr 2140

Db 6701 CCTTCTCCAGAGTCTTTTCTGGTGGACGGAGTCAGATCCATAGGTTTGCCCCCATTA 6760
Qy ProlyseProPheArgAspGluValSerPheCysValGlyLeuAenSerPheValVal 2160
Db 6761 CCGAAGCCGTTTTCGGGATGAGGTCTCGTTCTGCGTTCGAGTAAATTCATTTCTGCTG 6820
Qy GlySerGlnLeuProCysAspProGluProAspThrAspValLeuMetSerMetLeuThr 2180
Db 6821 GGGTCTCAGCTCCCTTGGATCCTGAACCTGACACAGCTATTGACGTCCATGCTAACA 6880
Qy AspProSerHisIleThrAlaGluThrAlaAlaArgArgLeuAlaArgGlySerProPro 2200
Db 6881 GACCCATCCCATATACCGCGCGAGACTCGACGGCGGTTTGCAACGGGGGTCAACCCCG 6940
Qy SerGluAspSerSerAlaSerGlnLeuSerAlaProSerLeuArgAlaThrCysThr 2220
Db 6941 TCCGAGGCAAGCTCTCAGCGAGCCAGCTATCGGCACCATCGCTCGAGCCACCTGCACC 7000
Qy ThrHisGlyLysAlaTyAspValAspMetValAspAlaAsnLeuPheMetGlyLysAsp 2240
Db 7001 ACCCAAGGCAAGCTCTATGATGTGACATGATGTGATGCCAACCTGTTATGGGGGGCGAT 7060
Qy ValThrArgIleGluSerGlySerLysValValValLeuAspSerLeuAspProMetVal 2260
Db 7061 GTACCCGATAGTCTGAGTCCAAAGTGGTCTGACTCTCTCGACCCCAATGGTC 7120
Qy GluLysSerAspLeuGluProSerIleProSerGluTyMetLeuProLysLysArg 2280
Db 7121 GAAGAAAGGCGACCTTGAGCCTTCGATACCATCGGAATATATGCTCCCAAGAGAGA 7180
Qy PheProAlaLeuProAlaTTPAlaArgProAspTyAsnProLeuValGluSer 2300
Db 7181 TTCCCAACAGCCCTTACCGGCTTGGCAAGGCTGATATCAACCCACCGCTTGTGGATCG 7240
Qy TrpLysArgProAspTyGlnProAlaThrValAlaGlyCysAlaLeuProProArg 2320
Db 7241 TGGAGGCGGAGATATCAACCGGCGCCTGTTGGGGGCTGGGCTCTCCCGCCCGCTAAG 7300
Qy LysThrProThrProProArgArgArgThrValGlyLysSerGluAspSerIle 2340
Db 7301 AAAACCCGACCGCTCCCGCAAGCAGACGCGGACAGTGGGTCTGAGTGAGAGCTCCATA 7360
Qy GlyAspAlaLeuGlnGlnLeuAlaIleLysSerPheGlyGlnProProSerGlyAsp 2360
Db 7361 GCAGATGCCCTACACAGCTGGCCATCAAGTCTTTGGCCAGCCCGCCCGGCGGAT 7420
Qy SerGlyLeuSerThrGlyAlaGlyAlaAlaAspSerGlySerGlnThrProProAspGlu 2380
Db 7421 TCAGGCTTTTCACGGGGGCGGACGACGCGATTCGGGAGTTCGGACCGCCCGCGATGAG 7480
Qy LeuAlaLeuSerGluThrGlySerIleSerSerMetProProLeuGluGlyLysLeuGly 2400
Db 7481 TTGGGCTTTTCGAGACAGGTTCCATCTCTCCATGCCCTCTCGAGGGGAGGCTGGA 7540
Qy AspProAspLeuGluProGluGlnValGluProGlnProProGlnGlyValAla 2420
Db 7541 GATCCAGACTTGGAGCTGAGCAGGTAGACTTCAACCTCCCGCCAGGGGGGGTGGTA 7600
Qy AlaProGlySerAspSerGlySerTrpSerThrCysSerGluLysAspSerValVal 2440
Db 7601 ACCCGGCTCAGGCTCGGGGTCTGGTCTACTTGTCTCGAGGAGGACACTCCGCTG 7660
Qy CysSerMetSerTySerTrpThrGlyAlaLeuIleThrProCysSerProGluGlu 2460
Db 7661 TGCTGCTCATGATATCTCTGACCGGGGCTTAATACTCTTGTAGCCCGGAGAG 7720
Qy GluLysLeuProIleAsnProLeuSerAsnSerLeuAlaGlyTyHisAsnLysValTy 2480
Db 7721 GAAAGTTGGCAATTGGCCCTTGAGCACTCTCTGTTGGATATATCAACAAGGCTGAT 7780
Qy CysThrThrThrLysSerAlaSerLeuArgAlaLysValThrPheAspArgMetGln 2500

Db 7781 TGTACCAATCAAGAGCGCTCATTAAAGGCTAAAAAGGTAACTTTTGTATAGCATCAA 7840
Qy ValLeuAspSerTyTyAspSerValLeuLysAspIleLysLeuAlaAlaSerLysVal 2520
Db 7841 GGGCTCGACGCTCATATAGCTCACTTCTTGAAGGACATTAAGCTAGCGGCTCTCCAGGTC 7900
Qy ThrAlaArgLeuLeuThrMetGluLysAlaCysGlnLeuThrProProHisSerAlaArg 2540
Db 7901 ACCGCAAGGCTTCTCACTTAGAGGAGGCTTCCAGTTAACTCCACCCCACTCTCTGCAAGA 7960
Qy SerLysTyGlyPheGlyAlaLysGluValArgSerLeuSerGlyArgAlaValAsnHis 2560
Db 7961 TCCAAGTATGGTGTGGGCTAAGGAGGTCCGACGCTTGTCCGGAGAGCGCTTAACCA 8020
Qy IleLysSerValTrpLysAspLeuLeuGluAspSerLeuThrProIleProThrThrIle 2580
Db 8021 ATCAAGTCCGTGTGGAGGACCTCTCTGGAGACACACAAACCCATTCCTTACAACTC 8080
Qy MetAlaLysAsnGluValPheCysValAspProThrLysGlyLysLysAlaAlaArg 2600
Db 8081 ATGCCCAAAATGAGGTGTTCTGCTGGACCCCAACAAAGGGGGGTAAAGAACGAGCTCGC 8140
Qy LeuIleValTyProAspLeuGlyValArgValCysGluLysMetAlaLeuTyAspIle 2620
Db 8141 CTATATCGTTTACCTCGACCTCGCGCTCAGGGTCTCGGAGAAAATGGGCTTTATGATATC 8200
Qy ThrGlnLysLeuProGlnAlaValMetGlyAlaSerTyGlyPheGlnTySerProAla 2640
Db 8201 ACACAAAGCTTCTCTCAGCGGCTGATGGGGCTTCTTATGATTCAGTACTCTCCCGCT 8260
Qy GluArgValGluPheLeuLeuLysAlaTrpAlaGluLysLysAspProMetGlyPheSer 2660
Db 8261 CAGCGGTGGAGTTCCTTGAAGGATGGCGGAAAGAACCCCTATGGGGTTTTTCG 8320
Qy TyrAspThrArgCysPheAspSerThrValThrGluArgAspIleArgThrGluGluSer 2680
Db 8321 TATGATACCGGATGCTTTGACTCAACCGCTCACTGAGAGACATCAGSACTGAGGAGTCC 8380
Qy IleTyArgAlaCysSerLeuProGluLysAlaHisThrAlaIleHisSerLeuThrGlu 2700
Db 8381 ATATATCGGCTTGTCTTCTTGGCCGAGGAGGCCACACTGCCATACACTCACTGACTGAG 8440
Qy ArgLeuTyValGlyGlyProMetPheAsnSerLysGlyGlnThrCysGlyTyArgArg 2720
Db 8441 AGACTTTACGTGGAGGGGCCCATGTTCAACAGAACGGCCAGACTCGGGGTACAGGCGT 8500
Qy CysArgAlaSerGlyValLeuThrThrSerMetGlyAsnThrIleThrCysTyValLys 2740
Db 8501 TGCGCGCCAGCGGGGTGCTTACCACTAGCATGGGAAACACCATCATCTATGTGAAA 8560
Qy AlaLeuAlaAlaCysLysAlaAlaGlyIleAlaProThrMetLeuValCysGlyAsp 2760
Db 8561 GCCTTAGCGGCTGTAAAGGCTGACGGGATAATTTGGCCCAATGCTGTGTATCGGCGAT 8620
Qy AspLeuValValIleSerGluSerGlnTyThrGluGluAspGluArgAsnLeuArgAla 2780
Db 8621 GACTTGTGTCTCATCTCAGAGAGCCAGGGGACCGAGGAGGAGCGGAACTTGAGAGCC 8680
Qy PheThrGluAlaMetThrArgTySerAlaProProGlyAspProProArgProGluTy 2800
Db 8681 TTCAAGAGGCTATGACAGGTATTTCTGCTCTCTCTGCTGACCCCGCCAGACCGAATAT 8740
Qy AspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaLeuGlyProGlnGly 2820
Db 8741 GACTTAGGCTGATAAATCTTGTCTCTCAATGTGTCTGTGGGCTTGGGCCCAAGGC 8800
Qy ArgArgArgTyTyThrLeuThrArgAspProThrThrProIleAlaAlaAlaTrpGlu 2840
Db 8801 CGCGCAGATACTTACTGACAGAGACCTTACCACTCCATCGGCCCGGGGTGCTCTGGA 8860
Qy ThrValArgHisSerProValAsnSerTrpLeuGlyAsnIleIleGlnTyAlaProThr 2860
Db 8861 ACAGTTAGACACTTCCCTGTCAATTCATGGCTAGGAAATCATCATCAGTACGCCCAACC 8920

Claim 5; Page 13-25; 36pp; Japanese.
 The invention relates to a human polypeptide related to hepatitis C virus (HCV), and the polynucleotide encoding it. The polypeptide can be used for the development of gene therapy on fulminant hepatitis C. This sequence represents cDNA encoding a human HCV-related polypeptide.
 Sequence 9678 BP; 1948 A; 2915 C; 2723 G; 2092 T; 0 other;

Alignment Scores:					
Pred. No.:	0	Length:	9678		
Score:	14943.00	Matches:	2765		
Percent Similarity:	95.42%	Conservative:	129		
Best Local Similarity:	91.16%	Mismatches:	139		
Query Match:	92.70%	Indels:	0		
DB:	24	Gaps:	0		

US-09-980-559-2 (1-3033) x ABK88904 (1-9678)

Qy	1	MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnAspArgProGln	20
CC	341	ATGAGCAAAATCCTAAACCTCAAGAAATAACCAAAAGAAACACCAACCGTCGCCCAAGAA	400
Qy	21	AspValLysPheProGlyGlyGlyGlnIleValGlyGlyValLysLeuLeuProArgArg	40
DB	401	GACGTTAAGTTCCTCGCGCGCGCGCCAGATCGTTGCGGAGTATATCTTGTTGCCCGCGCAGG	460
Qy	41	GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly	60
DB	461	GGCCCCAGGTGTGGTGTGCGCAGCAAGGAAACTTCGAGCGCTCCAGCCACAGCTGGG	520
Qy	61	ArgArgGlnProIleProLysAspArgArgSerThrGlyLysSerTrpGlyLysProGly	80
DB	521	AGACGCCAGCCATCCCTCAAGATCGGCGCTCCACTGCGAAGGCTGGGGAAACCAAGT	580
Qy	81	TyrProTrpProLeuTyrGlyAsnGluGlyLeuGlyTrpAlaGlyTrpLeuLeuSerPro	100
DB	581	CGCCCCGTGGCCCCCTATATGGGAATAGGAGACTGGCTGGGAGGATGCTCTGTGTCCCC	640
Qy	101	ArgGlySerArgProSerTrpGlyProAsnAspProArgHisArgSerArgAsnValGly	120
DB	641	CGAGGCTCTCGCCCCCTCTCTGGGGCCCCCATGACGCCCGCGCATAGCTCGCGCAACGTGGT	700
Qy	121	LysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProValVal	140
DB	701	AAAGTCATCGACACCCTAAAGTGTGGCTTTGCGGACCTCATGTGGGTATATCCCCGTGTA	760
Qy	141	GlyValaProLeuGlyGlyValaAlaArgAlaLeuAlaHisGlyValaArgValaLeuGluAsp	160
DB	761	GGCGCCCCGTAGTGGCGCCCGCCAGAGCTGTGCGCGCACGGGTGAGAGTCTCTGAGGAC	820
Qy	161	GlyValaAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeuAla	180
DB	821	GGGGTAAATATGCAACAGGGAACCTACCCGGTTTCCCTTTCTATCTTCTCTGCTGCC	880
Qy	181	LeuLeuSerCysIleThrThrProValSerAlaAlaGluValLysAsnIleSerThrGly	200
DB	881	CTGTTGTCTCGCATACCGTTCCGCTCTCTGTCGCCAGGTGCAAGAAATACCAAGTAGCAGC	940
Qy	201	TyrMetValThrAsnAspCysThrAsnAspSerIleThrTrpGlnLeuGlnAlaAlaVal	220
DB	941	TACATGTTGACCAATCACTGCTCCAATCACAGCATCACTTGGCAGCTCGAGGCTCGCGTT	1000
Qy	221	LeuHisValProGlyCysValProCysGluLysValGlyValAsnAlaSerGlnCysTrpIle	240
DB	1001	CTCAAGCTCCCCGGGTGGTCTTCGTGCGGAGAGTGGGGGATACGTCACGGGTGTTGGGT	1060
Qy	241	ProValSerProAsnValaAlaValGlnArgProGlyAlaLeuThrGlnGlyLeuArgThr	260
DB	1061	CCAGTCTCGCAAAACATGCTGTGGCGGACCGCGCGCTTACGACAGGGTCTCGCAGC	1120
Qy	261	HisIleAspMetValValMetSerAlaThrLeuCysSerAlaLeuTyrValGlyAspLeu	280

ORGANISM Hepatitis C virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.
 REFERENCE 1 (bases 1 to 9589)
 AUTHORS Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A., Miyakawa, Y. and Mayumi, M.
 TITLE Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions
 J. Gen. Virol. 72 (Pt 11), 2697-2704 (1991)
 JOURNAL MEDLINE 9204440
 PUBMED 1658196
 REFERENCE 2 (sites)
 AUTHORS Han, J. H. and Houghton, M.
 TITLE Group specific sequences and conserved secondary structures at the 3' end of HCV genome and its implication for viral replication
 Nucleic Acids Res. 20 (13), 3520 (1992)
 JOURNAL MEDLINE 92335016
 PUBMED 1321416
 REFERENCE 3 (sites)
 AUTHORS Hotta, H., Doi, H., Hayashi, T., Furukawa, M., Soemarto, M., Mizokami, M., Ohta, K. and Homma, M.
 TITLE Analysis of the core and E1 envelope region sequences of a novel variant of hepatitis C virus obtained in Indonesia
 Arch. Virol. 136 (1-2), 53-62 (1994)
 JOURNAL MEDLINE 94270990
 PUBMED 7545932
 COMMENT These data kindly submitted in computer readable form by: Hiroaki Okamoto
 Immunology Division
 Jichi Medical School
 Kawachi-gun
 Tochigi-Ken 329-04
 Japan
 Phone: 0285-44-2111 x3334
 Fax: 0285-44-1557.
 Location/Qualifiers
 1. 9589
 /organism="Hepatitis C virus"
 /mol_type="genomic RNA"
 /isolate="HC-J6"
 /db_xref="taxon:11103"
 341..9442
 /codon_start=1
 /product="polyprotein"
 /protein_id="BA00792.1"
 /db_xref="GI:21651"
 /translation="MSTNPKPQRKTKRNTNRPDQVFPFGGQIVGGVYLLPRGRRLGVRAKTKRSOPRRQPIPKRRSTGKSGKPGYPWPLXGNEGLWAGWLLSPRGSRPSGPNDRHRNRVGVKVIDLTCGPDADLMGYIPVVGAPLGVAALAHGVLELGVNATGNDLPGCSFIFILLALSCITTPVSAEYKNTSTGYMTNDSTWQLOAVLHVPCCVPCVKVGTSCRCWIPVSNVAVQPGALTOGLRTHDMVMSATLCSALYVGLCGVMIAAQMFIVSPQHHNFVQDCNCSIYPGTITGHRMAMDMWMSPTAMILAVAMRPVPIIDIIIGAHVGVNFGVLVPSMOGAWKVVVILLAAQVDAQTHVGSITAHNARTLGMFSLGAROKIOLINTNGSHINRTALCNDLSLHTGLASLPTHSNNSGCPERSACRSIEAFVNGALQYEDVTHPEDMRKYCHRYPRQGVVSASVCGPVYCTPSRVVGTDLRGLPPTVNGENEDVFLNSTRPFGCSWFCCTWNSSTYTKTGAPGCRIRADNASMDLLCPTDCRKHPTDIYKCGSGPMLTFLCLIDIPYRLMHPCTTUNYTIPLKMTVGVGSRHLTAACNFTRGDKNCNDRSLSPLHSTTEWAILPCTYSDLPATGLLHLHONIIVDQFMGLSPALTKYIWRWVWVLLFLLADARVCACLWMLILQAALAEKAYDLHASVHGOI GAALLVILPTLTGKYTLRSFLMLCYLLTLSPGLLALPOQAYDASVHGOI GAALLVILPTLTGKYTLRSFLMLCYLLTLAEAWQEWAPMVGQGRDGIIMAVAI PCFQVFDITKMLLAVLAGPAYLLKGLTRVPYFPAHALRMTCTVYHLAGGRVYVQVLLALGRWTGTIYDHLTPMSMAANGRLDLAVAVEPIIFSPMKYIVMGAETAAACGDI LHLGLPVSARLGRVLLGADGYTSKGSLLAPITAYAQTRGLLGTIVSMTHGRDKTEQAGEI QVLSVTQSPFLGTTISGLVLTWYHAGNKTLAGSRPVTOMYSSAEGDLVGVPSPGTKSLPECTCGAVDLXLYNRADVIPARRGRGALLSPRLSTLKGSSGVPVLCPRHAGVGFRAVYSGRVAKSIDPFPVETLDIVTRSPVSDNSTPPAVPOTYOVGVLLHATPGSKTKVPVAVAAQGVKVLINLPSVLDLPGFAYLKGANGININETGVTVTTCGATYSTYKELADGCGAYDIIICDECHAVDSITLIGTVLQDOASTAGVLTATATATPGSVTTTHPHNEEVALGQGEIPIFYCBALPISLYIKGSHLIFCHSKKCDLALALNGMLNAVATIRGLDVSIVPTQGDVWVATDALMTFTGDFDSVIDCNVAVTQVVDFSLDPTFTITITQTVFQDAVSRQRGR

TCGRLGIYRYVSTGCRASGMPDSVVLCECDAGANMYELTPTAETTVRLRAVNTPTPL
 PVQCRLFEWEAVFTGLTHIDHFLSTQKSGENFAYLTATQVACARAKAPPPSDV
 MCKLRLKPTLTVGLTLLYRLGTSVTEVLTHTPVTKYIATQMOADEVMTVTVLAG
 GLAAAVAKICQGLAQKQASQADIQAPVQASWPKVQFQWAKMFWNFISIGITLAGIST
 LPGNPAVMAFSAALTSPLSTTILMLIGLWASQIAPAGATGVPVSPALGVGA
 VSGILGKVLVDLAGYAGISGALVAFKMSGKPSMEDVNNLLPGILLSPALGVGA
 ICAALIRRHVGPGEAGVMNRLIAFASRGHNVAPHYTVTESDASORVQLGLSTIT
 SLILRLHNNITEDCFIPCSGSLRWLDVMDVTCITLIDPKNLTSLKPKMPGLPFTSCQ
 KGKVGWAGTGIMTRPCGANISGNVLSGMRITGPKTKMNTWQTEPIKCYTSGQC
 VPKPAPNPKIALIRVAASEYAEVTHGSHYITGLTDLNKKVPCOLPSPSPSWDGV
 QHREAPIPKPFREDFVSCVGLANFPVGSOLPCDEPDDVLTSLTSPSHITAE
 ARELARGPPESEASSASOLSPSLRATCTHGAANDVMDVNDLVNPPGDSVTRISES
 KVVLDSLDLPMVEERSDLPSIPSEYMLPKRPPALPAPWAPDIPAPPDVEVTRIS
 QYTVAGCGLPPPKTIPPTPPRRRTVGLSESSIAUKLOALIKSFGQPPPGDGLS
 TGAADSGSRTPPDELATSETGISISSPPEGEPPGDDLEPEQVBLQPPQGVVPT
 GSGSGMSTCEEDSVVCCSMYSWTGALITPCSPSEKLPINPLNSLLARHNKY
 CITTSASIRAKKVTDFDMQALDAHYDKLAAASKVTRALLLEAQCILTPHYS
 ARSKYFGAKBVRSLGSAVNHKSVKDLLEDITQPIPTTMAKNEVECVDTKGGK
 KAARLIVPDLGVRCEMALYDITQKLPQAVMGASYGFOYSPAQRVFLKAMAEK
 DPMFSDYTRCFDSTVTERDIRTEESYIRACSLPEBAHTIHSLTRVYCGPMSK
 GOTCYRRCRASGLVITSMGNTITCYVKAACAAGIITAPMLVCGDLDLVVSSQG
 TEBEDNRLRATPEAMTRYSPAPGPPRPDYDLITSCSSNVSVAGPQGRRTYLR
 DPTTIRAAWETVHRSVSNWLCNIIIOYAPTIWARMVLMTHFFSLMAQDRLDONL
 FEMYGAVYSVSPDLPLPAIIRLHGLDAFSLRTHYTHLTVASALRKLGLAPLRWKS
 RARVASLISRGGRAAVCGVLPFNWAVTKLKLTPLEARNLLDLSWFTVAGGGDI
 YHSVSRAPRRLILGLLLLVGVGLFLLPAR"
 BASE COUNT 1970 a 2820 c 2633 g 2166 t
 ORIGIN
 Query Match 92.4%; Score 9973.6; DB 14; Length 9589;
 Best Local Similarity 96.0%; Pred. No. 0;
 Matches 9204; Conservative 0; Mismatches 384; Indels 0; Gaps 0;
 QY 1 ACCCGGCCCTAATAGCGCGACACTCCGCGCATGAATCACTCCCTGTGAGGAAGTACTGT 60
 DB 1 ACCCGGCCCTAATAGCGCGACACTCCGCGCATGAATCACTCCCTGTGAGGAAGTACTGT 60
 QY 61 CTTACAGCAGAAAGCGTCTAGCGCATGCTGTGAGTGTCTGTACAGCTCCAGGCC 120
 DB 61 CTTACAGCAGAAAGCGTCTAGCGCATGCTGTGAGTGTCTGTACAGCTCCAGGCC 120
 QY 121 CCGCTTCCCGGAGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGGAATTGCGGG 180
 DB 121 CCGCTTCCCGGAGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGGAATTGCGGG 180
 QY 181 AAGACTGGGTCTTTCTTGGATAAACCACTCTATGCCCGGCATTTGGCGCTGCCCCCG 240
 DB 181 AAGACTGGGTCTTTCTTGGATAAACCACTCTATGCCCGGCATTTGGCGCTGCCCCCG 240
 QY 241 CAAGACTGCTAGCGGAGTAGCGTTGGGTTCGGAAGCGCTTGTGCTACTGCTGATAGG 300
 DB 241 CAAGACTGCTAGCGGAGTAGCGTTGGGTTCGGAAGCGCTTGTGCTACTGCTGATAGG 300
 QY 301 TCGTTGCGAGTCCCGGGAGGTCTCTAGACCGTGCACCAAGAGCTTAACTTTCCGGGG 360
 DB 301 TCGTTGCGAGTCCCGGGAGGTCTCTAGACCGTGCACCAAGAGCTTAACTTTCCGGGG 360
 QY 361 TCAGAGAAAAACAAAAAGAAACCAACCGTCCGCCCAAGAGCTTAACTTTCCGGGG 420
 DB 361 TCAGAGAAAAACAAAAAGAAACCAACCGTCCGCCCAAGAGCTTAACTTTCCGGGG 420
 QY 421 CCGCCAGATCGTTGGCGGAGTATCTTGTGTCGCGCGACGGGGCCAGGTGCGGTGCG 480
 DB 421 CCGCCAGATCGTTGGCGGAGTATCTTGTGTCGCGCGACGGGGCCAGGTGCGGTGCG 480
 QY 481 CGCGACAGGAGACTTTCGAGCGGTCCAGCCACTGCGAGCGCGCCAGCCCATCCCTAA 540
 DB 481 CGCGACAGGAGACTTTCGAGCGGTCCAGCCACTGCGAGCGCGCCAGCCCATCCCTAA 540
 QY 541 AGATCGCGCTCCACTCGCAAAATCTCGGGGAAAAACAGGATACCCCTCGCCCTATACGG 600
 DB 541 CGATCGCGCTCCACTCGCGAAATCTCGGGGAAAAACAGGATACCCCTCGCCCTATACGG 600

8581 TGCAGGGGTAATCGCCGACCAATGCTGTATGCGGGATGACTTGGTTGTCACTCAGA 8640
8641 AACCCAGGGGACCGAGGAGGACGAGGAACTTGAGAGCCTTCAACGAGGCTATGACAG 8700
8641 AACCCAGGGGACCGAGGAGGACGAGGAACTTGAGAGCCTTCAACGAGGCTATGACAG 8700
8701 GTATTCTGCCCCCTCTCTGCTGAGCCCCCCCCCAGACCGGAGTATGATCTCGAGCTGATACATC 8760
8701 GTATTCTGCCCCCTCTCTGCTGAGCCCCCCCCCAGACCGGAGTATGATCTCGAGCTGATACATC 8760
8761 TTGCTCTCAAAATGTCGTGCGGCTGGGCCCCCAGAGCGCGCCGAGATACTACTGAC 8820
8761 TTGCTCTCAAAATGTCGTGCGGCTGGGCCCCCAGAGCGCGCCGAGATACTACTGAC 8820
8821 CAGAGACCTTACACTTCOAATGCGCGGCTGGGAAACAGTTAGACACTCCCTCTGT 8880
8821 CAGAGACCTTACACTTCOAATGCGCGGCTGGGAAACAGTTAGACACTCCCTCTGT 8880
8881 CAATTCAATGCTGGGAAACATCATCCAGTAGCGGCCCGACCATATGSGCTGCGATGCTCT 8940
8881 CAATTCAATGCTGGGAAACATCATCCAGTAGCGGCCCGACCATATGSGCTGCGATGCTCT 8940
8941 GATGACACACTTCTCTCAATCTCATGGCTCAGACAGCGCTGACCAAGCTCAACTT 9000
8941 GATGACACACTTCTCTCAATCTCATGGCTCAGACAGCGCTGACCAAGCTCAACTT 9000
9001 TGAGATGACGAGGCGGTGACTCCGTTGAGTCCCTTGAGCTCCAGCTATATGAAAG 9060
9001 TGAGATGACGAGGCGGTGACTCCGTTGAGTCCCTTGAGCTCCAGCTATATGAAAG 9060
9061 GTTACATGCGCTGTGACGCTTTTCTCTGACACATACACTCCCGACCAAGCTGACAGCGGT 9120
9061 GTTACATGCGCTGTGACGCTTTTCTCTGACACATACACTCCCGACCAAGCTGACAGCGGT 9120
9121 GCGTTACGCGCTCAGAAACTTGGGCGCCACCTCTCAGAGCTGAGAGCGCGGACG 9180
9121 GCGTTACGCGCTCAGAAACTTGGGCGCCACCTCTCAGAGCTGAGAGCGCGGACG 9180
9181 TGCAGTCAAGGCGTCCCTCACTCTCGTGGGAGAGCGCGCTTGGCTGATATCT 9240
9181 TGCAGTCAAGGCGTCCCTCACTCTCGTGGGAGAGCGCGCTTGGCTGATATCT 9240
9241 CTTCAATTGGGCGGTGAGAGCAAGCTCAAACTCACTCCATTTGCGGAGCGCGCTCT 9300
9241 CTTCAATTGGGCGGTGAGAGCAAGCTCAAACTCACTCCATTTGCGGAGCGCGCTCT 9300
9301 GGAATTTATCCAGCTGCTTCACTCGTGGGCGCGCGCGGCGGACATTTATCAGCGGTCT 9360
9301 GGAATTTATCCAGCTGCTTCACTCGTGGGCGCGCGCGGCGGACATTTATCAGCGGTCT 9360
9361 GCGTCCCGGACCGCGCTTATGCTCTTTGGCTTACTCTTCTTTTGTAGGGGTAGGCT 9420
9361 GCGTCCCGGACCGCGCTTATGCTCTTTGGCTTACTCTTCTTTTGTAGGGGTAGGCT 9420
9421 TTTCCTACTCCCGCTCGGTAGAGCGGCAACAATTAGCTTACACTCCATAGCTTACCTCC 9480
9421 TTTCCTACTCCCGCTCGGTAGAGCGGCAACAATTAGCTTACACTCCATAGCTTACCTCC 9480
9481 CTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9540
9481 CTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9540
9541 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9600
9541 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 9600
9601 TACTTCTTTCTTTGCTTCACTTACCGCTTACTGACAGCTAGCTGTGAAAGGTCCT 9660
9601 TACTTCTTTCTTTGCTTCACTTACCGCTTACTGACAGCTAGCTGTGAAAGGTCCT 9660
9661 GAGCGGATGACTGACAGAGTGCCTTAATCTGCTCTCTCAGATCATGT 9711

9661 GAGCGGATGACTGACAGAGTGCCTTAATCTGCTCTCAGATCATGT 9711

RESULT 3
AAF23486
ID AAF23486 standard; DNA, 9711 BP.
XX
AC AAF23486;
XX
DT 21-MAR-2001 (first entry)
XX
DE Infectious Hepatitis C virus 2a genotype.
XX
KW GBV-B; hepatitis C virus; HCV; vaccine; ds.
XX
OS Hepatitis C virus.
XX
PN WO200075337-A1.
XX
PD 14-DEC-2000.
XX
PF 02-JUN-2000; 2000WO-US15293.
XX
PR 04-JUN-1999; 99US-0137694.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Bukh J, Yanagi M, Emerson SU, Purcell RH;
XX
DR WPI; 2001-091214/10.
XX
PT New infectious nucleic acids of the GB virus-B clone, useful for
PT indirectly studying the molecular properties of hepatitis C virus (HCV)
PT and in developing vaccines and therapeutics for HCV -
XX
PS Disclosure; Page 78-82; 96pp; English.
XX
CC The present invention relates to GB virus-B. The nucleic acid molecules
CC of the invention are useful for indirectly studying the molecular
CC properties of hepatitis C virus (HCV). The infectious nucleic acid
CC sequence of the GB virus-B clone and the HCV/GBV-B chimera may be used
CC in the development of vaccines and therapeutics for HCV.
XX
SQ Sequence 9711 BP; 1989 A; 3851 C; 2668 G; 2203 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 9707.8; DB 22; Length 9711;
Matches 9709; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACCGCGCCCTAATAGGGGCGACACTCCGCAATGAATCACTCCCTGTGAGGAACTACTGT 60
DB 1 ACCGCGCCCTAATAGGGGCGACACTCCGCAATGAATCACTCCCTGTGAGGAACTACTGT 60
QY 61 CTTCAAGCAGAAAGGCTCTAGCCATGGGTTAGTATGAGTGTGTACAGCTCCAGGCC 120
DB 61 CTTCAAGCAGAAAGGCTCTAGCCATGGGTTAGTATGAGTGTGTACAGGCCCTCCAGGCC 120
QY 121 CCCCCTCCCGGAGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGAATGCGGG 180
DB 121 CCCCCTCCCGGAGAGCCATAGTGTCTGCGGAACCGGTGAGTACACCGAATGCGGG 180
QY 181 AGAGCTGGGTCTTTCTTGGATAAACCCACTCTATGCGCGGCCATTTGGCGGTGCCCCCG 240
DB 181 AGAGCTGGGTCTTTCTTGGATAAACCCACTCTATGCGCGGCCATTTGGCGGTGCCCCCG 240
QY 241 CAAGACTGTAGCCAGTAGGCTTGGGTTGCGAAAGGCTTGTGGTACTGCTGTAGGG 300
DB 241 CAAGACTGTAGCCAGTAGGCTTGGGTTGCGAAAGGCTTGTGGTACTGCTGTAGGG 300
QY 301 TCGTTGCGAGTGCCTCCCGGAGGTCTCCGTAGCCGTGACCGCTGACCAATCTTAACC 360
DB 301 TCGTTGCGAGTGCCTCCCGGAGGTCTCCGTAGCCGTGACCGCTGACCAATCTTAACC 360

QY	8041	CCTCTGGAGACTCAGAAACCA	CTATTCCCAACCAACCTTATGCCC	AAAAATGAGGTGT	8100	
DB	8041	CCTCTGGAGACTCAGAAACCA	CTATTCCCAACCAACCTTATGCCC	AAAAATGAGGTGT	8100	
QY	8101	CTCGTGGACCCCAACCAAGGGGGG	CAGAAAGCAGCTCGCCCTTATCG	TTTACCCTCGACCT	8160	
DB	8101	CTCGTGGACCCCAACCAAGGGGGG	CAGAAAGCAGCTCGCCCTTATCG	TTTACCCTCGACCT	8160	
QY	8161	CGCGCTCAGGGTCTCGCGAGAGAT	GGCCCTTTATGACATTACACAA	AAACCTTCTCAGGC	8220	
DB	8161	CGCGCTCAGGGTCTCGCGAGAGAT	GGCCCTTTATGACATTACACAA	AAACCTTCTCAGGC	8220	
QY	8221	GGTCATGGGGGCTCTTATGGAAT	CCAGTATCCCGCGCTCAGCGGG	TAGAGTTCTCTT	8280	
DB	8221	GGTCATGGGGGCTCTTATGGAAT	CCAGTATCCCGCGCTCAGCGGG	TAGAGTTCTCTT	8280	
QY	8281	GAAGACATGGCGCGAAGAGAGCA	CCCTATGGGTTTTTGGTATGAT	ACCGATGCTTTGA	8340	
DB	8281	GAAGACATGGCGCGAAGAGAGCA	CCCTATGGGTTTTTGGTATGAT	ACCGATGCTTTGA	8340	
QY	8341	CTCAACCGTCTCTGAGAGAGACAT	CAGGACTGAGGAGTCCATATAT	CGGGCCCTGCTCTT	8400	
DB	8341	CTCAACCGTCTCTGAGAGAGACAT	CAGGACTGAGGAGTCCATATAT	CGGGCCCTGCTCTT	8400	
QY	8401	GCCCGAGGGGCCCACTGCGCAT	ACACTCCCTAACTGAGAGACTT	TAGTGGAGGGCC	8460	
DB	8401	GCCCGAGGGGCCCACTGCGCAT	ACACTCCCTAACTGAGAGACTT	TAGTGGAGGGCC	8460	
QY	8461	TATGTTCAACGACAGGGCCAA	ACCTGCGGCTACAGCGGTTG	CGGCCAGCGGGTGCT	8520	
DB	8461	TATGTTCAACGACAGGGCCAA	ACCTGCGGCTACAGCGGTTG	CGGCCAGCGGGTGCT	8520	
QY	8521	CACCACTAGCATGGGGAA	CAACCATCATGCTACGTGA	AAAGCCTTAGCGGCTTGTAAAGC	8580	
DB	8521	CACCACTAGCATGGGGAA	CAACCATCATGCTACGTGA	AAAGCCTTAGCGGCTTGTAAAGC	8580	
QY	8581	TGCAGGGATTAATCGCGCCCA	CAATGCTGGTATGCGGCGAT	GACTTGGTTGTCACTCAGA	8640	
DB	8581	TGCAGGGATTAATCGCGCCCA	CAATGCTGGTATGCGGCGAT	GACTTGGTTGTCACTCAGA	8640	
QY	8641	AAGCCAGGGGACCGAGGAGGAC	GCAGCGAACCTGAGAGCCTT	CACGAGGCTATGACCAG	8700	
DB	8641	AAGCCAGGGGACCGAGGAGGAC	GCAGCGAACCTGAGAGCCTT	CACGAGGCTATGACCAG	8700	
QY	8701	GTAATCTGCCCTCTCTGGTGA	CCCCCGACGAGATGATCTGG	AGCTGATTAAGTC	8760	
DB	8701	GTAATCTGCCCTCTCTGGTGA	CCCCCGACGAGATGATCTGG	AGCTGATTAAGTC	8760	
QY	8761	TTGCTCTCAAAATGTTCTGTGG	CGCTTGGGCCACAAAGCGCG	CGCAGATACTCTGAC	8820	
DB	8761	TTGCTCTCAAAATGTTCTGTGG	CGCTTGGGCCACAAAGCGCG	CGCAGATACTCTGAC	8820	
QY	8821	CAGAGA	CCCTACCACTCGCCGGGCTG	CTGGGAAACAGTTAGACACT	CCCCCTGT	8880
DB	8821	CAGAGA	CCCTACCACTCGCCGGGCTG	CTGGGAAACAGTTAGACACT	CCCCCTGT	8880
QY	8881	CAATTATGGCTGGGAAACAT	CATCCTCAGTAGCGCCGAC	CAATAGGCTCGCATGGTCT	8940	
DB	8881	CAATTATGGCTGGGAAACAT	CATCCTCAGTAGCGCCGAC	CAATAGGCTCGCATGGTCT	8940	
QY	8941	GATGACACTTCTCTCAATCT	CTATGGGCTCAGACAGGCT	GGAACGAACCTCACTT	9000	
DB	8941	GATGACACTTCTCTCAATCT	CTATGGGCTCAGACAGGCT	GGAACGAACCTCACTT	9000	
QY	9001	TGAGATGATACGAGCGGTG	TACTCTGGAGTCCCTTGG	ACCTCCAGCTATAATTTGAAAG	9060	
DB	9001	TGAGATGATACGAGCGGTG	TACTCTGGAGTCCCTTGG	ACCTCCAGCTATAATTTGAAAG	9060	
QY	9061	GTTTACATGGGCTTGACGCT	TTTTTCTCTGCAACATACA	CTCCCCACGAACCTGACACGGGT	9120	
DB	9061	GTTTACATGGGCTTGACGCT	TTTTTCTCTGCAACATACA	CTCCCCACGAACCTGACACGGGT	9120	

QY 9121 GCGTCAGCCCTCAGAAACTTGGGGGCGCCACCCCTCAGAGCGTGGAGAGCGCGGCAG 9180
 DB 9121 GCGTCAGCCCTCAGAAACTTGGGGGCGCCACCCCTCAGAGCGTGGAGAGCGCGGCAG 9180
 QY 9181 TGCAGTCAGGGCGTCCCTCATCTCCGTCGGGGGAGAGCGCGCTTTCGGTGGATATCT 9240
 DB 9181 TGCAGTCAGGGCGTCCCTCATCTCCGTCGGGGGAGAGCGCGCTTTCGGTGGATATCT 9240
 QY 9241 CTTCAATTGGGCGGTGAAGAGCAAGAGCTCAAACTCACTCCATTGCGGAGCGCGCTCCT 9300
 DB 9241 CTTCAATTGGGCGGTGAAGAGCAAGAGCTCAAACTCACTCCATTGCGGAGCGCGCTCCT 9300
 QY 9301 GGATTTATCCAGCTGTTTACCGTCGGCGCGGGGGGCGACATTATATCAGAGGTGTC 9360
 DB 9301 GGATTTATCCAGCTGTTTACCGTCGGCGCGGGGGGCGACATTATATCAGAGGTGTC 9360
 QY 9361 GCGTCCCGACCCCGCTTATTCCTTTTGGCTACTCTACTTTTGTAGGGGTAGGCT 9420
 DB 9361 GCGTCCCGACCCCGCTTATTCCTTTTGGCTACTCTACTTTTGTAGGGGTAGGCT 9420
 QY 9421 TTTCTTACTCCCGCTCGGTAGAGCGGCACACATTAGCTACACTCCATAGCTAACTGCC 9480
 DB 9421 TTTCTTACTCCCGCTCGGTAGAGCGGCACACATTAGCTACACTCCATAGCTAACTGCC 9480
 QY 9481 CTTT 9540
 DB 9481 CTTT 9540
 QY 9541 TT 9600
 DB 9541 TT 9600
 QY 9601 TACTTTCTTTCTGTGTGCTCATCTTAGCCCTAGTACAGCGCTAGCTGTGAAGTCCGT 9660
 DB 9601 TACTTTCTTTCTGTGTGCTCATCTTAGCCCTAGTACAGCGCTAGCTGTGAAGTCCGT 9660
 QY 9661 GAGCCGATGACTGACAGAGGTGCGTAACCTGCTCTCTGACAGATCATGT 9711
 DB 9661 GAGCCGATGACTGACAGAGGTGCGTAACCTGCTCTCTGACAGATCATGT 9711

RESULT 4
 ID AAQ38218 standard; cDNA to mRNA; 9589 BP.
 AC AAQ38218;
 XX
 DT 25-MAR-2003 (updated)
 DT 01-JUL-1993 (first entry)
 DE NAMBH virus strain HC-J6 genome cDNA sequence.
 XX
 XX Non A non B hepatitis virus; amplification; HC-J1; HC-J8; plasma; ss.
 XX
 XX Non A, non B hepatitis virus strain HC-J6.
 PH Key Location/Qualifiers
 FT CDS 341..9442
 FT /*tag= a
 XX
 PN EP532167-A2.
 PD 17-MAR-1993.
 PF 30-JUL-1992; 928P-0306952.
 PR 09-AUG-1991; 91JP-0287402.
 PR 05-DEC-1991; 91JP-0360441.
 PA (IMMO) IMMUNO JAPAN INC.
 XX Nakamura T, Okamoto H;
 XX

DR WPI: 1993-087166/11.
 DR N-PSDB; AAR33538.
 PT Polynucleotide(s), polypeptide(s) and antibodies of NAMBH virus -
 PT useful for detecting NAMBH, as a vaccine and for screening blood
 XX samples
 XX Claim 2; Page 27-32; 93pp; English.
 XX RNA was isolated from the plasma of human patients positive for
 CC NAMBH virus (strain HC-J6) and was subjected to reverse transcription
 CC to produce cDNA. The resulting cDNA was amplified by PCR, and
 CC nucleic acid sequences determined by analysis of both clones from the
 CC cDNA library and clones obtd. by PCR amplification (36 clones in total).
 CC The NAMBH HC-J6 genome was found to contain an open reading frame
 CC encoding a polypeptide precursor of 3033 amino acid residues.
 CC See also AAQ38172-221.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX
 XX Sequence 9589 BP; 1968 A; 2820 C; 2635 G; 2166 T; 0 other;
 SQ
 Query Match 92.4%; Score 8970.4; DB 14; Length 9589;
 Best Local Similarity 96.0%; Pred. No. 0;
 Matches 9202; Conservative 0; Mismatches 386; Indels 0; Gaps 0;
 QY 1 ACCCGCCCTTAATAGGGGCGACACTCCGCAATGAATCACTCCCTGTGAGGAACTACTGT 60
 DB 1 ACCCGCCCTTAATAGGGGCGACACTCCGCAATGAATCACTCCCTGTGAGGAACTACTGT 60
 QY 61 CTTCAAGCAGAAAGCGTCTAGCCATGAGTGTAGTATGATGCTGTACAGCTCCAGGCC 120
 DB 61 CTTCAAGCAGAAAGCGTCTAGCCATGAGTGTAGTATGATGCTGTACAGCTCCAGGCC 120
 QY 121 CCCCTCCCGGAGAGCCTAGTGTCTCGGAAACCGGTGAGTACACCGAATTCGCGG 180
 DB 121 CCCCTCCCGGAGAGCCTAGTGTCTCGGAAACCGGTGAGTACACCGAATTCGCGG 180
 QY 181 AAGACTGGTCTCTTTCTTGATAAACCCACTCTATGTCGCGCAATTTGGGGGTGCCCCCG 240
 DB 181 AAGACTGGTCTCTTTCTTGATAAACCCACTCTATGTCGCGCAATTTGGGGGTGCCCCCG 240
 QY 241 CAAGACTGTAGCCGAGTACGTTGGTTTGCAGAAAGGCTTGTGTACTGTCTGATAGG 300
 DB 241 CAAGACTGTAGCCGAGTACGTTGGTTTGCAGAAAGGCTTGTGTACTGTCTGATAGG 300
 QY 301 TGTCTGGAGTGTGCGGAGGTCTGTAGACGCTGACGTCAGCAGCAATCTTAACC 360
 DB 301 TGTCTGGAGTGTGCGGAGGTCTGTAGACGCTGACGTCAGCAGCAATCTTAACC 360
 QY 361 TCAAGAGAAAACCAAAAGAAACCAACCAACCGTCCCAACAGACGTTAAGTTTCCGGCGG 420
 DB 361 TCAAGAGAAAACCAAAAGAAACCAACCAACCGTCCCAACAGACGTTAAGTTTCCGGCGG 420
 QY 421 CGGCCAGATCGTTGGCGAGTATCTTGTTCGCGCAGGGGCGCCAGTTGGGTGTGG 480
 DB 421 CGGCCAGATCGTTGGCGAGTATCTTGTTCGCGCAGGGGCGCCAGTTGGGTGTGG 480
 QY 481 CGCGAAGAGAGACTTTCGAGCGGTCCAGCAGAGTGTGAGGCGCCAGCCATCTTA 540
 DB 481 CGCGAAGAGAGACTTTCGAGCGGTCCAGCAGAGTGTGAGGCGCCAGCCATCTTA 540
 QY 541 AGATCGCGCTCCACTGGCAAAATCTTGGGGAACCAAGGATACCTTGGCCCTTATACGG 600
 DB 541 AGATCGCGCTCCACTGGCAAAATCTTGGGGAACCAAGGATACCTTGGCCCTTATACGG 600
 QY 601 GAATGAGCGACTCGGCTGGGCGAGGATGGCTCTGTTCCTCCCGAGGTTCCTGCTCTTG 660
 DB 601 GAATGAGCGACTCGGCTGGGCGAGGATGGCTCTGTTCCTCCCGAGGTTCCTGCTCTTG 660
 QY 661 GGGGCCCCAATGACCCCGCATAGGTCCCGCAACGTTGGGTAGGTGTCATGATACCTTAAC 720
 DB 661 GGGGCCCCAATGACCCCGCATAGGTCCCGCAACGTTGGGTAGGTGTCATGATACCTTAAC 720

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2004, 06:04:40 ; Search time 13743 Seconds
(without alignments)
11773.505 Million cell updates/sec

Title: US-09-980-559-1
Perfect score: 9711
Sequence: 1 accgcgccctaataggggcg.....ggtctctcgcagatcatgt 9711

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_da.*

2: gb_hg.*

3: gb_in.*

4: gb_ov.*

5: gb_ov.*

6: gb_ph.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_ro.*

12: gb_sy.*

13: gb_sy.*

14: gb_un.*

15: gb_vl.*

16: gb_vl.*

17: em_hum.*

18: em_hum.*

19: em_mu.*

20: em_mu.*

21: em_or.*

22: em_or.*

23: em_ov.*

24: em_ov.*

25: em_ph.*

26: em_ph.*

27: em_ro.*

28: em_ro.*

29: em_un.*

30: em_un.*

31: em_vl.*

32: em_vl.*

33: em_hg_hum.*

34: em_hg_hum.*

35: em_hg_inv.*

36: em_hg_inv.*

37: em_hg_mus.*

38: em_hg_mus.*

39: em_hg_pln.*

40: em_hg_pln.*

41: em_hg_rod.*

42: em_hg_rod.*

43: em_hg_mam.*

44: em_hg_mam.*

45: em_hg_vrt.*

46: em_hg_vrt.*

47: em_hg_hum.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9711	100.0	9711	6	AX057086 Sequence
2	9711	100.0	9711	6	AX057317 Sequence
3	9711	100.0	9711	6	AX057395 Sequence
4	9711	100.0	9711	14	AF177036 Sequence
5	8973.6	92.4	9589	14	HPCPOLP
6	8970.4	92.4	9589	6	E07361
7	8970.4	92.4	9589	6	E07362
8	8970.4	92.4	9589	6	E07362
9	8507.4	87.6	9659	14	AB047641
10	8425.8	86.8	9700	14	AF169005
11	8370.2	86.2	9693	14	AF169003
12	8367.8	86.2	9653	14	AF169004
13	8364.2	86.1	9683	14	AB047645
14	8349	86.0	9691	14	AB047644
15	8329.4	85.8	9661	14	AF169002
16	8273.4	85.2	9674	14	AB047640
17	8233.4	84.8	9678	14	AB047642
18	8210.2	84.5	9416	14	AF238481
19	8186.8	84.3	9416	14	AF238482
20	8169.2	84.1	9416	14	AF238483
21	8156.4	84.0	9416	14	AF238484
22	8126.4	83.7	9416	14	AF238485
23	8093.4	83.2	9641	14	AB047643
24	8020.6	82.6	9678	6	BD160781
25	8020.6	82.6	9678	14	AB047639
26	6479.2	66.7	9513	14	D50409
27	6383.6	65.7	9488	14	AB031663
28	6048	62.3	9654	14	AB030907
29	6031.4	62.1	9511	6	E07468
30	6031.4	62.1	9511	6	E07468
31	6027	62.1	9511	6	E07467
32	5933.8	61.7	9511	14	HPC36G
33	5936	61.1	9416	14	AF238486
34	5816.8	59.9	9611	14	AF177037
35	5800.8	59.7	9611	6	AX057088
36	5800.8	59.7	9611	6	AX057092
37	5800.8	59.7	9611	14	AF177039
38	5692	58.6	9611	6	AX057090
39	5692	58.6	9611	14	AF177038
40	5676	58.4	9611	6	AX057094
41	5676	58.4	9611	14	AF177040
42	4625.6	47.6	9646	6	AB110828
43	4625.6	47.6	9646	6	BD069982
44	4625.6	47.6	9646	14	AF009606
45	4620.8	47.6	9646	12	AF387806

ALIGNMENTS

RESULT 1	AX057086	Sequence 1 from Patent WO0075338.	9711 bp	DNA	linear	PAT 17-JAN-2001
LOCUS	AX057086	Sequence 1 from Patent WO0075338.				
DEFINITION	AX057086	Sequence 1 from Patent WO0075338.				
ACCESSION	AX057086	Sequence 1 from Patent WO0075338.				
VERSION	AX057086.1	GI:12309919				
KEYWORDS						
SOURCE	Hepatitis C virus					
ORGANISM	Hepatitis C virus					
	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;					
REFERENCE	1					
AUTHORS	Yanagi M., Bukh J., Emerson S.U. and Purcell R.H.					
TITLE	Cloned genome of infectious hepatitis C virus of genotype 2a and uses thereof					

Pred. No. is the number of results predicted by chance to have a

ORGANISM Hepatitis C virus
 Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus
 1 (bases 1 to 9589)
 Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A.,
 Miyakawa, Y., and Mayumi, M.
 Nucleotide sequence of the genomic RNA of hepatitis C virus
 isolated from a human carrier: comparison with reported isolates
 for conserved and divergent regions
 J. Gen. Virol. 72 (Pt 11), 2697-2704 (1991)
 92044440
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 HAN, J. H. and Houghton, M.
 2 (sites)
 1658196
 Nucleic Acids Res. 20 (13), 3520 (1992)
 Group specific sequences and conserved secondary structures at the
 3' end of HCV genome and its implication for viral replication
 Nucleic Acids Res. 20 (13), 3520 (1992)
 92335016
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 Hotta, H., Doi, H., Hayashi, T., Purwanta, M., Soemarto, W.,
 Mizokami, M., Ohba, K. and Honma, M.
 Analysis of the core and E1 envelope region sequences of a novel
 variant of hepatitis C virus obtained in Indonesia
 Arch. Virol. 136 (1-2), 53-62 (1994)
 94270990
 MEDLINE
 PUBMED
 REFERENCE
 COMMENT
 These data kindly submitted in computer readable form by: Hiroaki
 Okamoto
 Immunology Division
 Jichi Medical School
 Kawachi-gun
 Tochigi-ken 329-04
 Japan
 Phone:
 Fax:
 0285-44-2111 x3334
 0285-44-1557.
 Location/Qualifiers
 1. 9589
 /organism="Hepatitis C virus"
 /mol_type="genomic RNA"
 /isolate="HC-J6"
 /db_xref="taxon:11103"
 341. 9442
 /codon_start=1
 /product="polyprotein"
 /protein_id="BA00792.1"
 /db_xref="GI:221651"
 /translation="MSTNPKPQKTKRNTNRPDQVKFGGQIVGGVYLLPRGPRIL
 GVRATKTSERSOPRRQPIPKDRRTSGKSGKPGYPWPPLYGNEGLWAGWLLSPRG
 SRPSMGNDPRHSRNKGVDTLTCCGADLMGYIPVVGAPLGGVARALAHGVRVLED
 GYNFATGNLPGCSFSIFLLALSCITTPVSAABVKRLSTGYMTNDCTNDSITWOLQA
 AVLVPGVPCPKVGTNRKCIIPVSNVAVQCPGALTQGLRTHIDMVMSATLCSALY
 VGLCCGVMLAAQMTIVSPQHHMVFQDCNCSLYEGTITGHRNWDMMNWSFTATNLL
 AVAMRVPVYIDYIGAGKGMFGLAFPSMQQANAKVYVILLAAQVDHQTHVCGGST
 ANHARTLCMPLSGARQKLIINTNGSHWLNRLNLCNDSLTGFLASLFTYHSFNSS
 GCFERMSACRSLEAFVWGALQYEDNVTNEDRPPYCHMHPYPRQCGVVSASSVCGPV
 YCTFSPVVGTTDRLGATYTWGENETDVLNSTRPFGSGWFGCTWNSSTGYTKTC
 GAPPCIRDFNASHMDLLCPTDCRKHPTDYYIKCGSGPWLTPCLIDYPRLWHPYC
 TWNYTIFKRTMYGVGHRHTAACNFTGRDRCNLEDRDSQLSPLLHSITTEMAILPCT
 YSDLPALSTGLLHLQNIIVQVFMGLSPALTKYIVRWKVVLLFLLADARVCACLW
 MLTLQAEALAEKLVLRASASACNGFLYFVIFVAAWYIKGRVVPATVYSYTLGLW
 SFLGLLALPQAVAYDASVHQIGAAVLVLTITLTPGKNTLSRFLMCLYLLTL
 AEMVQENAPPMQVGGSDGIIIMAVAFPCPGVVDITKMLLVAGPVLVIGNLTTPV
 YFVRHALLRMEKTVIHLAGGRYQWVLLALGRWSTYLYDHLTPMSWANGRLDLA
 VAVEPIIFSPMEKTVIHWAGTAACGDIHLGLPVSARLGRVILGPALGYSKGWSLL
 APITAYAGTGLLGLTIVSWTGDHTEQAGEIQLYSTVQSFLGTISGVLMVYVHG
 AGNKLASRGPTQKMSAGDLYGMFSPGKLSLEPCTCGAVDLIVTRNADYIPA
 RRGDKRGLSLPFLSLTKSGSGVPLCPGRHAGVFRFAAVCSRGVAKSIDFIPVET
 LQIVTSPFTSDNSTRPVPAYQVYGLHAPTSGSKTKVPVAYAAQGTQKVLNPSV
 AATLGFATLSKAGHNINRTVTVTGTAPYSTYKGLADGGCAGGAYDIICD
 ECHAVDSTTLIGTVLQDQETAGVRLTVLATATPPGSVTTPPHNIBSVLQSGEIP
 FYGRAIPUSYKGGHLLFCHSKKXCDLALRGNLNAVATYTRGLDVSIVPTQGDV
 VVATDALMTGTGDFSDVIDCNVAVTVQVDFSLDPTPTIT"QTVPQDAVRSORRRG

CDS

FEATURES
 source

1603

W 00075

BASE COUNT 1970 a 2820 c 2633 g 2166 t

ORIGIN

Query Match 92.4%; Score 8973.6; DB 14; Length 9589;
 Best Local Similarity 96.0%; Pred. No. 0;
 Matches 9204; Conservative 0; Mismatches 384; Indels 0; Gaps 0;

Oy 1 ACCCGCCCTAATAGCGGACACTCGCCCATGATCACTCCCTCTGAGGAACACTACTGT 60
 Db 1 ACCCGCCCTAATAGCGGACACTCGCCCATGATCACTCCCTCTGAGGAACACTACTGT 60

Oy 61 CTTTCAGCAGAAAGCGTCTAGCCATGCGGTAGTATGAGTGTCTGATACACCTCCAGGCC 120
 Db 61 CTTTCAGCAGAAAGCGTCTAGCCATGCGGTAGTATGAGTGTCTGATACACCTCCAGGCC 120

Oy 121 CCCCCTCCCGGAGAGCCATAGTGGTCTGCGGAACCGGTGAGTACACCGGAATTCGCGGG 180
 Db 121 CCCCCTCCCGGAGAGCCATAGTGGTCTGCGGAACCGGTGAGTACACCGGAATTCGCGGG 180

Oy 181 AAGACTGGGTCTCTTCTTGGATAAACCCACTATATCCCGGCCNTTGGCGGTGCGCCCG 240
 Db 181 AAGACTGGGTCTCTTCTTGGATAAACCCACTATATCCCGGCCNTTGGCGGTGCGCCCG 240

Oy 241 CAAGACTCTAGCCGAGTAGCGTGGGTTCGGAAGCGCTCTGTGTTACTGCTGATAGGG 300
 Db 241 CAAGACTCTAGCCGAGTAGCGTGGGTTCGGAAGCGCTCTGTGTTACTGCTGATAGGG 300

Oy 301 TCGTTTCGAGTGCCCGGAGGTCTCTGATAGCGTGCACCATGAGCACAATACTTAAACC 360
 Db 301 TCGTTTCGAGTGCCCGGAGGTCTCTGATAGCGTGCACCATGAGCACAATACTTAAACC 360

Oy 361 TCAGAGAAAAACCAAGAAACCAACCTCTGCGCCACAGAGCTTAACTTTCGCGGG 420
 Db 361 TCAGAGAAAAACCAAGAAACCAACCTCTGCGCCACAGAGCTTAACTTTCGCGGG 420

Oy 421 CGGCGCAGATCTGTGCGGAGTATATCTTGTTCGCGCAGGGGCCCGAGTTGGGTGCGG 480
 Db 421 CGGCGCAGATCTGTGCGGAGTATATCTTGTTCGCGCAGGGGCCCGAGTTGGGTGCGG 480

Oy 481 CGCGCAAGAGGAGACTTCGAGCGGTCCAGGCCACCTGGAAGCGCGCAGCCCATCTCTAA 540
 Db 481 CGCGCAAGAGGAGACTTCGAGCGGTCCAGGCCACCTGGAAGCGCGCAGCCCATCTCTAA 540

Oy 541 AGATCGGCGCTCCTCGCAAACTCTCGGGAACACAGGATACCTCTGCGCCCTATAGG 600
 Db 541 AGATCGGCGCTCCTCGCAAACTCTCGGGAACACAGGATACCTCTGCGCCCTATAGG 600

Db 9221 GCGGTTTCGGCGGATATCTTCACTGGCGGTGAGCCAGCTCAACTCACTCA 9280
 Qy 2981 LeuProGluAlaArgLeuLeuAspLeuSerSerTrpPheThrValGlyAlaGlyGly 3000
 Db 9281 TTCCCGGAAGCGCGCTCTGATATTATCAGCTGGTTCACTGTCCGCGCGCGGGG 9340
 Qy 3001 AspIleThrHisSerValSerArgAlaArgProArgLeuLeuPheGlyLeuLeu 3020
 Db 9341 GACATTATACAGAGGTGTCCGTCGCCGCGCGCTTATTACTCTTGGCTACTCTTA 9400
 Qy 3021 LeuPheValGlyValGlyLeuPheLeuLeuProAlaArg 3033
 Db 9401 CTTTTGTAGGGGTAGGCTTTTCTACTTCCCGCTCGG 9439

RESULT 8
 112861
 LOCUS 112861
 DEFINITION Sequence 2 from patent US 5428145.
 ACCESSION 112861
 VERSION 112861.1 GI:910242
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED.
 REFERENCES 1 (bases 1 to 9589)
 AUTHORS Okamoto,H. and Nakamura,T.
 TITLE Non-A, non-B, hepatitis virus genome, polynucleotides,
 polypeptides, antigen, antibody and detection systems.
 JOURNAL Patent: US 5428145-A 2 27-JUN-1995;
 FEATURES
 source Location/Qualifiers
 1..9589
 /organism="unknown"
 BASE COUNT 1968 a 2820 c 2635 g 2166 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0 Length: 9589
 Score: 15824.00 Matches: 2966
 Percent Similarity: 98.78% Conservative: 30
 Best Local Similarity: 97.79% Mismatches: 37
 Query Match: 98.16% Indels: 0
 DB: 6 Gaps: 0

US-09-980-559-2 (1-3033) x 112861 (1-9589)

Qy 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
 Db 341 ATGAGCAACAATCTTAACCTCAAGAGAAACCAAGAAACCAACCAACCGTCCGCCA 400
 Qy 21 AspValLysPheProGlyGlyGlnIleValGlyValGlyValLysLeuProArgArg 40
 Db 401 GAGCTTAACTTCGGCGCGCGGCAGATCGTTGGCGGAGTAACTTGTTCGGCGCAGG 460
 Qy 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 60
 Db 461 GCGCCGAGGTGGGTGGCGCGCGACAGAGAGACTTCGAGCGGTCCAGCCACAGTGA 520
 Qy 61 ArgArgGlnProIleProLysAspArgSerThrGlyLysSerTrpGlyLysProGly 80
 Db 521 AGCGCCAGCCCATCTCCATAGGATCGCGCTCCACTGGCAATCTCGCGGAAACCAAGA 580
 Qy 81 TyrProTrpProLeuThrGlyAsnGluGlyLeuGlyTrpAlaGlyTrpLeuLeuSerPro 100
 Db 581 TACCCCTTGGCCCTATACGGGATGAGGACTCGGCTGGCGAGATGCTCTCTTCTCCCC 640
 Qy 101 ArgGlySerArgProSerTrpGlyProAsnAspProArgHisArgSerArgAsnValGly 120
 Db 641 CGAGGTTCCTGCTCTTGGCGGCCCAATGACCCCGCATAGGTCCCGCAACGCGGT 700
 Qy 121 LysValIleAspThrLeuThrCysGlyPheAlaAspLeuMetGlyTyrIleProValVal 140
 Db 701 AAGGTCATCATGATCCCTAAAGTCGGGCTTTGCCGACCTCATGGGGTACATCCCTGTGTA 760

Qy 141 GlyAlaProLeuGlyGlyValAlaArgAlaLeuAlaHisGlyValArgValLeuGluAsp 160
 Db 761 GCGCCCGCTTCGCGCGCGCTGCCAGAGCTCTCGCGCATGGCGTGAGAGTCTCTGGAGC 820
 Qy 161 GlyValAsnPheAlaThrGlyAsnLeuProGlyCysSerPheSerIlePheLeuLeuAla 180
 Db 821 GGGGTAAATTTGCAACAGGAACCTTACCCTGGTCTCTCTCTCTCTCTCTCTCTCTCT 880
 Qy 181 LeuLeuSerCysIleThrThrProValSerAlaGluValLysAsnIleSerThrGly 200
 Db 881 CTGCTGTCTGTGATCACACCCCGCTCTCCGCTCCGAGTGAAGACATCATGATCCGCG 940
 Qy 201 TyrMetValThrAsnAspCysThrAsnAspSerIleThrTrpGlnLeuGlnAlaVal 220
 Db 941 TACATGTGACCAACCACTGCACCAATAGATAGCTTACCTGGCACTCCAGGCTGCTGC 1000
 Qy 221 LeuHisValProGlyCysValProCysGluLysValGlyAsnAlaSerGlnCysTrpIle 240
 Db 1001 CTCCACGCTCCCGCTGCTGCTGCGAGAAAGTGGGGAATACATCTCGTGTCTGATA 1060
 Qy 241 ProValSerProAsnValAlaValGlnArgProGlyAlaLeuThrGlnGlyLeuArgThr 260
 Db 1061 CCGGTCTCACCGAATGTGCGGTGCAGCAGCCCGCGGCTCTCACCGAGGCTTACGAGC 1120
 Qy 261 HisIleAspMetValValMetSerAlaThrLeuCysSerAlaLeuThrValGlyAspLeu 280
 Db 1121 CACATTGACATGTTGTGATGTCCGCCAGCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 1180
 Qy 281 CysGlyGlyValMetLeuAlaAlaGlnMetPheIleValSerProGlnHisHisTrpPhe 300
 Db 1181 TCGGTGGCTGTGATGCTTGCAGCCAGATGTTCTATGTTCTCGCACAGCACCACTGGTT 1240
 Qy 301 ValGlnAspCysAsnCysSerIleThrProGlyThrIleThrGlyHisArgMetAlaTrp 320
 Db 1241 GTGCAAGACTGCATTTGCTTCACTTACCTGTGTACCATCACTGCAGACCGCATGGCTG 1300
 Qy 321 AspMetMetAsnTrpSerProThrAlaThrMetIleLeuAlaThrAlaMetArgVal 340
 Db 1301 GACATGATGATGACTGTGTCGCCAGCTACCATGATCTTCGGCTGCGATCGCGGTC 1360
 Qy 341 ProGluValIleAspIleSerGlyAlaHisTrpGlyValMetPheGlyLeuAla 360
 Db 1361 CCGAGGTTCATCATAGACATCATTTGGCGGGCTCATTTGGGGCGCTCATGTTGGGCTTAG 1420
 Qy 361 TyrPheSerMetGlnGlyAlaTrpAlaLysValValIleLeuLeuLeuAlaAlaGly 380
 Db 1421 TACTTCTCTATGCGAGCGGTGGCAAAAGTGTGTGTCATTTCTTTTCTGCTGGCGCGG 1480
 Qy 381 ValAspAlaArgThrHisThrValGlyGlySerAlaAlaGlnThrThrGlyArgLeuThr 400
 Db 1481 GTGGACCGGCAACCCATACCGTGTGGGGTCTACCGCGCATACCGCAGAGACCTCACC 1540
 Qy 401 SerLeuPheAspMetGlyProArgGlnLysIleGlnLeuValAsnThrAsnGlySerTrp 420
 Db 1541 GGCATGTTCTCTCTGCTGCGCGAGGAGAAATCCAGCTCATCAACCAATGCGAGTGG 1600
 Qy 421 HisIleAsnArgThrAlaLeuAsnCysAsnAspSerLeuHisThrGlyPheIleAlaSer 440
 Db 1601 CACATCAACCGCACCCCTCTGAATGCAATGACTCTTTGTCACACCGGCTCTCTCGCTCA 1660
 Qy 441 LeuPheThrHisSerPheAsnSerSerGlyCysProGluArgMetSerAlaCysArg 460
 Db 1661 CTGTTCTACCCACAGACTTCACTCTGTCAGAGTGTCCGAGCAATGTCCGCTGCGCG 1720
 Qy 461 SerIleGluAlaPheArgValGlyTrpGlyAlaLeuGlnThrGluAspAsnValThrAsn 480
 Db 1721 AGTATCGAGGCTTTTGGGTGGGATGGGGCGCTTTACAAATATAGGACAAATGTCACCA 1780
 Qy 481 ProGluAspMetArgProTrpCysTrpHisTrpProProArgGlnCysGlyValValSer 500
 Db 1781 CAGAGGATATGAGACCGTATTGCTGGCACTACCCACCAAGACAGTGTGCTGTAGTCTCC 1840

292- aa level

8141 CTTATCGTTTACCTGACCTCGGCTCAGGCTCGCGAGAAATGCGCCCTTTATGATATC 8200
 2621 ThrGlnLeuProGlnAlaValMetGlyAlaSerTyrGlyPheGlnTyrSerProAla 2640
 8301 ACAAAAGCTTCTCAGCGGTGATGGGGCTCTTTATGGAATTCAGTACTCCCGGCT 8260
 2641 GlnArgValGluPheLeuLeuLysAlaTTPAlaGluLysLysAspProMetGlyPheSer 2660
 8361 CAGCGGTGAGGTTTCTCTTGAAGGATGGCGGAAAGAACACCTATGGTTTTCG 8320
 2661 TyrAspThrArgCysPheAspSerThrValThrGluArgAspIleArgThrGluGluSer 2680
 8321 TATGATACCGATGCTTTCATCAACCTCTCACTGAGAGACATCAGGACTCAGAGATCC 8380
 2681 IleTyrArgAlaCysSerLeuProGluAlaHisThrAlaIleHisSerLeuThrGlu 2700
 8381 ATATATCGGGCTTGTCTTCCGCCGAGGAGCCACACTGCCATACACTCACTGCTGAG 8440
 2701 ArgLeuTyrValGlyGlyProMetPheAsnSerLysGlyGlnThrCysGlyTyrArgArg 2720
 8441 AGACTTTAGCTGGAGGGCCCATGTTCAACGACGAGGCCAGACCTGCGGGTACAGGCT 8500
 2721 CysArgAlaSerGlyValLeuThrThrSerMetGlyAsnThrIleThrCysTyrValLys 2740
 8501 TGGCGCGCCAGCGGGGTCTTACCACTAGCATGGGACACCATCATGCTATGTGAA 8560
 2741 AlaLeuAlaCysGlyAlaAlaGlyIleIleAlaProThrMetLeuValCysGlyAsp 2760
 8561 GCCTTAGCGGCTGTAAAGCTGCAAGGATTAATTCGCCGCCAATGCTGTATGCGGCGAT 8620
 2761 AspLeuValIleSerGluSerGlnGlyThrGluAspGluArgAsnLeuArgAla 2780
 8621 GACTGTGTTGTCATCTCAGAGAGCCAGCGGACCGAGGAGGACGAGCGGAACCTGAGAGCC 8680
 2781 PheThrGluAlaMetThrArgTyrSerAlaProGlyAspProProArgProGluTyr 2800
 8681 TTCAGGAGGCTATGACCAAGTATCTGCGCCCTCTGCTGAGCCGCCAGACCGGAAAT 8740
 2801 AspLeuGluLeuIleThrSerCysSerSerAsnValSerValAlaLeuGlyProGlnGly 2820
 8741 GACTCGAGCTGATTAACATCTTCTCTCTCAAAATGCTGTGCGGTGGGCCCAACAGGC 8800
 2821 ArgArgTyrTyrTyrLeuThrArgAspProThrThrProIleAlaArgAlaAlaTyrGlu 2840
 8801 CGCGCAGATACCTACCTACGAGAGCCCTACCACTCCAATCGCGGCTGCGCTCGCGAA 8860
 2841 ThrValArgHisSerProValAsnSerTyrLeuGlyAsnIleIleGlnTyrAlaProThr 2860
 8861 ACAGTTAGACTCCCTGTCTCATTTCTGCTAGGAAACATCATCGTACGCCCCCAACC 8920
 2861 IleTyrAlaArgMetValLeuMetThrHisPheSerIleLeuMetAlaGlnAspThr 2880
 8921 ATATGGCTCGCATGCTCTCATGACACATCTTCTTCCATTTCTATGCGCCCAAGTACT 8980
 2881 LeuAspGlnAsnLeuAsnPheGluMetTyrGlyAlaValTyrSerValSerProLeuAsp 2900
 8981 CTGGACAGAACCTCAACTTTGAGATGTAGCGAGCGGTGACTCTCGTGAATCCCTTGGAC 9040
 2901 LeuProAlaIleIleGluArgLeuHisGlyLeuAspAlaPheSerLeuHisThrTyrThr 2920
 9041 CTCCGAGCCATAATTGAAGGTATACCGGGCTTGACGCTTCTCTCTGACACATACACT 9100
 2921 ProHisGluLeuThrArgValAlaSerAlaLeuArgLysLeuGlyAlaProLeuArg 2940
 9101 CCCAGCAACTGACAGGGGTGCTTTCAGCCCTCAGAAAACCTTGGGGCGGCCACCTCTAGA 9160
 2941 AlaTyrLysSerArgAlaArgAlaValArgAlaSerLeuIleSerArgGlyValArgAla 2960
 9161 GGTGGAGAGAGCGGGACGTCAGTCAAGGCGCTCTCATCTCTCCCGTGGGGGAGAGCG 9220
 2961 AlaValCysGlyArgTyrLeuPheAsnTyrAlaValIleThrLysLeuLeuLeuThrPro 2980
 9221 GCGGTTGGCGCGATATCTCTTCACTTGGCGGGTGAAGACCACTCAACTCACTCTCA 9280

2981 LeuProGluAlaArgLeuLeuAspLeuSerSerTyrPheThrValGlyValaGlyGly 3000
 9281 TTGCGGAAGCGCGCTCTCTGATTTTCCAGCTGGTTCACTGTCCGCGCCCGCGGCG 9340
 3001 AspileTyrHisSerValSerArgAlaAspProArgLeuLeuPheGlyLeuLeuLeu 3020
 9341 GACATTATACAGCGGTGCTGCGTGGCGGACCCCTTATTACTCTCTGGCTACTCTTA 9400
 3021 LeuPheValGlyValGlyLeuPheLeuLeuProAlaArg 3033
 9401 CTTTGTGAGGGTACGCCCTTTTCTACTCCCGCGCTGG 9439
 RESULT 7
 E07362
 LOCUS CDNA of Hepatitis non-A non-B virus. linear PAT 29-SEP-1997
 DEFINITION E07362
 ACCESSION E07362
 VERSION E07362.1 GI:2175501
 KEYWORDS JP 1994121689-A/2.
 SOURCE unidentified
 ORGANISM unclassified.
 REFERENCE 1. (bases 1 to 9589)
 AUTHORS Okamoto, H. and Nakamura, T.
 TITLE NON-A NON-B HEPATITIC VIRUS GENE, POLYNUCLEOTIDE, POLYPEPTIDE,
 ANTIGEN AND ANTIBODY DETECTION SYSTEM
 JOURNAL Patent: JP 1994121689-A 2 06-MAY-1994.
 COMMENT OS Hepatitis non-A non-B virus
 PN JP 1994121689-A/2
 PD 06-MAY-1994
 PF 09-AUG-1991 JP 1991287402
 PI OKAMOTO HIROAKI, NAKAMURA TETSUO
 PC C12N15/51, A61K39/29, A61K39/395, A61K39/395, C07K13/00, C12P21/02,
 C12P21/08,
 PC G01N33/53, G01N33/576, G01N33/577//A61B10/00;
 CC strandedness: Single;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 FH Key Location/Qualifiers
 FT source 1. 9589
 /organism="Hepatitis non-A non-B virus" PT
 /strain="HCV-J6".
 FEATURES
 source Location/Qualifiers
 1. 9589
 /organism="unidentified"
 /mol_type="genomic RNA"
 /db_xref="taxon:32644"
 BASE COUNT 1968 a 2820 c 2635 g 2166 t
 ORIGIN
 Alignment Scores:
 Pred. NO.: 0 Length: 9589
 Score: 15824.00 Matches: 2966
 Percent Similarity: 98.78% Conservativity: 30
 Best Local Similarity: 97.79% Mismatches: 37
 Query Match: 98.16% Indels: 0
 DB: 6 Gaps: 0
 US-09-980-559-2 (1-3033) x E07362 (1-9589)
 QY 1 MetSerThrAsnProLysProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
 DB 341 ATGACCAATCTTAACCTCAAGAAACCAAAAGAAACCAACCGTCGCCCA 400
 QY 21 AspValLysPheProGlyGlyGlyGlnIleValGlyValTyrLeuLeuProArg 40
 DB 401 GACCTTAAGTTTCGGGGCGGCCAGATCGTTGGCGAGTATACTTGTTCGCCGCG 460
 QY 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 60

9281 TTGCGGAGGCGGCTCTCCGATTTATCCAGCTGGTTCACTGTCGGCGCGCGGCGGCG 9340
 3001 AspileTyHisSerValSerArgAlaArgProArgLeuLeuLeuPheGlyLeuLeu 3020
 9341 GACATTTATCAGAGGTGTCGGTCCGCGCGCGCGCTTATCTCTTGGCTTACTCTTA 9400
 3021 LeuPheValGlyValGlyLeuPheLeuLeuProAlaArg 3033
 9401 CTTTTTGTAGGCGGTAGGCGCTTTCTTCTACTCCCGCTCGG 9439

RESULT 6
 E07361
 LOCUS
 DEFINITION SRNA of Hepatitis non-A non-B virus, HCV-J6.
 ACCESSION E07361
 VERSION E07361.1 GI:5708558
 KEYWORDS JP 1994121689-A/1.
 SOURCE unidentified
 ORGANISM unidentified

REFERENCE
 1 (bases 1 to 9589)
 Okamoto, H. and Nakamura, T.
 NON-A NON-B HEPATITIS VIRUS GENE, POLYNUCLEOTIDE, POLYPEPTIDE,
 ANTIGEN AND ANTIBODY DETECTION SYSTEM
 Patent: JP 1994121689-A 1 06-MAY-1994;
 NAKAMURA TETSUO

COMMENT
 OS Hepatitis non-A non-B virus
 PN JP 1994121689-A/1
 PP 06-MAY-1994
 PF 09-AUG-1991 JP 1991287402
 PI OKAMOTO HIROAKI, NAKAMURA TETSUO
 FC C12N15/51, A61K39/29, A61K39/395, A61K39/395, C07K13/00, C12P21/02,
 C12P21/08,
 PC GOIN33/53, GOIN33/576, GOIN33/577//A61B10/00;
 CC strandedness: Single;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 FH Key Location/Qualifiers
 FT source 1..9589
 FT /organism="Hepatitis non-A non-B virus" FT
 FT /strain="HCV-J6".
 FEATURES
 source
 1..9589
 Location/Qualifiers
 /organism="unidentified"
 /mol_type="genomic RNA"
 /db_xref="taxon:32644"
 BASE COUNT 1968 a 2820 c 2635 g 2166 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0 Length: 9589
 Score: 15824.00 Matches: 2966
 Percent Similarity: 98.78% Conservative: 30
 Best Local Similarity: 97.79% Mismatches: 37
 Query Match: 98.16% Indels: 0
 DB: 6 Gaps: 0

US-09-980-559-2 (1-3033) x E07361 (1-9589)

Qy 1 MetSerThrAsnProGlyProGlnArgLysThrLysArgAsnThrAsnArgProGln 20
 Db 341 ATGAGCACAAATCTTAAACCTCAAGAGAAACCAACCAACCAACCAACCAACCAACCA 400
 Qy 21 AspValLysPheProGlyGlyGlnPheValGlyGlyValTyrlleuLeuProArgArg 40
 Db 401 GACGTTAAGTTTCGGCGCGCGCGAGATGTTGGCGGAGTACTTGTTCGGCGCGAGG 460
 Qy 41 GlyProArgLeuGlyValArgAlaThrArgLysThrSerGluArgSerGlnProArgGly 60
 Db 461 GCGCCCGAGGTTGGGTGTGGCGCGCAAGGAAGACTTCGAGCGGTCCCGAGCCAGTGA 520

2981 LeuProGluAaArgLeuLeuAspLeuSerSerTriphTherValGlyValaGlyGly 3000
 9281 TTCCGGAAGCGCGCTCTCGATTTATCCAGCTGCTTACCGTCCGCGCGCGCG 9340
 3001 AspleTyHiserValserArgAlaArgProArgLeuLeuPheGlyLeuLeu 3020
 9341 GACATTTATCACAGGCTGCTCGGTCGCGCGCGCGCGCTTATGCTCTTGCTACTCT 9400
 3021 LeuPheValGlyValGlyLeuPheLeuLeuProAlaArg 3033
 9401 CTTTGTGAGGCTGAGCGCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 9439

HPCPOLP 9589 bp RNA linear VRL 20-JUN-1998
 Hepatitis C virus genomic RNA for polyprotein, complete cds.
 D00944
 00944.1 G1:221650
 Polyprotein
 Hepatitis C virus
 Hepatitis C virus
 Hepatitis C virus
 Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae;
 Hepacivirus.
 1 (bases 1 to 9589)
 Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A.,
 Miyakawa, Y., and Mayumi, M.
 Nucleotide sequence of the genomic RNA of hepatitis C virus
 isolated from a human carrier: comparison with reported isolates
 for conserved and divergent regions
 J. Gen. Virol. 72 (Pt 11), 2697-2704 (1991)
 92044440
 1658196
 2 (sites)
 Han, J.H. and Houghton, M.
 Group specific sequences and conserved secondary structures at the
 3' end of HCV genome and its implication for viral replication
 Nucleic Acids Res. 20 (13), 3520 (1992)
 92335016
 1321416
 3 (sites)
 Hotta, H., Doi, H., Hayaishi, T., Purwanta, M., Soemarto, W.,
 Mizokami, M., Ohba, K. and Homma, M.
 Analysis of the core and E1 envelope region sequences of a novel
 variant of hepatitis C virus obtained in Indonesia
 Arch. Virol. 136 (1-2), 53-62 (1994)
 94270990
 7545932
 These data kindly submitted in computer readable form by: Hiroaki
 Okamoto
 Immunology Division
 Jichi Medical School
 Kawachi-gun
 Tochigi-ken 329-04
 Japan
 Phone: 0285-44-2111 x3334
 Fax: 0285-44-1557
 Location/Qualifiers
 1. 9589
 /organism="Hepatitis C virus"
 /mol_type="genomic RNA"
 /isolate="HC-J6"
 /db_xref="taxon:11103"
 341..9442
 /codon_start=1
 /product="polyprotein"
 /protein_id="BAA00792.1"
 /db_xref="GI:221651"
 /translation="MSTNPKPQRTKENTNRPPQDKVPGGQIVGGVYLLPRGPRRL
 GVRATKTSRSQPRGRROPIKDRRSTGSKGPGYPMPLPGNGLGWAGWLLSPRG
 SRPSGNDPDRSRNNGVKVITLTGCFADLMGYIPVGCAPLGVARALANGVRVLED
 GYNATGLPCSPSIFLLALLSCITTPVSAEVRNISTGVMTNDCTNDSITWLOA
 AVLVPGCVKGVNTSRWIPVSRVAVQDPGALTGCLTHIDMYNSATLCSALY
 VEDLCGGVLAQAQNPVSPQHMFVQDCNCSYPTTIGHRNANDMNNWSPFATNML

FEATURES
 source
 CDS

578

AYAMRVEVIIDIGGANGVMFGLAYFSMQGAWKVVILLAAAGVADQHTTVCST
 ANRTLTGMSFGARQIOLINTNGSHNRTALNCNDSHTGPTPLASITPTTHSPNS
 GPERASACRSLEAFVGVGALQEDVNTNEDNRPYCHYPPRQCVVSAVSQCPV
 YCFPTSPVVTGTTDLGATPTWGENEDVFLNSTRPPOGSMFGCTWNSCTYTKC
 GAPCIRIRADFNASMDLLCPTDCFRKHPDITIKGSGPMLTPRCCLIDVYRMLHYP
 TUNYTIKRVYGVGVEHRLTAACNFTGRDNCLEDRDRSOLSPHLLSTENAILPCT
 YSDLPALSTGLHLHQNIVQAPYMGSLSPALTKYIVRMWVLLFLLADARACVAM
 MTLTQABALEKLVLVAASASONGELYFVIFVAAKYIKRNVUPLATYSLGLW
 SFGLLLALPOQRYADSVNGQIGALLVLLTLTLTPYKTLISRFLLWLCYLLTL
 AEWQWAPQYVGRDGLIIVAFPCGVDFDITMGLAVGLAFGLKGLKTRP
 YFVRHALLRMCTVWHLAGGYVOMVLLALGRWTGTYIYDHLTPMSPDWAANGRLDLA
 VAVEPIIFSPMEKVIWGAETAACGDLHGLPVSARLREHVLGPGDSTYSGWMLL
 APITAAQGTGLLGTIVSMSTGRDTEQAGEIQVLSVTQAGVLYLTVNADVIPA
 AKNTLAGSGPVTQYSSAEGDLVGMSPPTKSEPTCCGADLYLTVNADVIPA
 RRGDGRGALLSPRLSTKSGSGPVLCPRHAGVGVRAAVSRGVAKSIDIPYEST
 LDITRSPFTSDNTPPAVPTQYGVGLHPTGSKTKVPAVAYAKIVLWVPSV
 AATLFGCAVLSKANGINIRTVITGAPITYSTYKGLADGAGCAGAYDIIICD
 ECHAVDSTTILGIGVLDQATAGVRLTVLATATPGSVTTPHNIEEVALGOGEP
 FYGRAIPLSYIKGRHLIFCHSKKKDELAALRGMLNAVYVRLGVDSVIRPQDGR
 VVATDALMTGTGDFSDVNDVAVTQVDFSLDPTFTLTITQVQDASVSQRGR
 TCEGRGLIYRVYVSTGERASGMPDSVLCGYDAGAAKYLTPATVRLAVPNTGL
 PVQDHLBFWAFTGLHIDARHLSQTKSGENFAYLTAQATVCABAKAPRSDV
 MKCLTLAPTIVGTFPLGLRGLSVINEVILTHPIKINATICQADLEWISTWLAG
 GYLAAYACLATGCVCIIGRLHVNRAVAPDKVEYEAPEDEBCASRAALISBGQ
 RIAEMUKSKIQLLOASKQADIDOPAVQASPKVEQFWACHMWNFISGLIYLGUST
 LPGNVAVSMAPSAALTPLSTSTILLNGLGLASQIAPAGATGFWVSLVGA
 VSGIGLVLDILAGYAGISGALVAFKMGSEKSMEDVNNLLFGIISPGALVVG
 ICALLRHHVGEAGVQVNRMLIAPASRGNHVAHPHYVTESDASQVTLGLSLTIT
 SCALRHHNITBCDPCIPGSGMLRDMVCTILTDFKMLTSLKPLMPGLPFTSCQ
 KGVYVAGTGMITRCPCGANISGNVRLSGMRITGPKTQNIWQGTFFPINCYTEQC
 VPKPAPFKIAIRVAASEYAEVTHQSGSYHITGLTDLNKLVCPLSPPEFSSWQGV
 OHRFAPIKPFRDBSVFCVLSVFSQSLPCDPRPDVLTSLMLTDPDSTHASTA
 ARELAGSPSSASASQSLASLATCTTHGKAYDVMDVMDANLFGWGDVTRIESG
 KVVLDLDMVEESDLESTPSEYMLPKKPPALPAWADPNPPLVESKRDY
 AVATGACALPPEKTPPTPBRRTVGLSESSADALQOLAIXKPGQPSGDSGLS
 TGADADSRTPPDELALSETGSISSNPPLGEGDPLEPEQVLELPQPPQGVVTR
 GSGSGWSTSEBDDSVCCSWSYMTGALITPCSPREKRLFINLSLRTHNRY
 CTTKSASLRKKVTFDRQADAHYDVLKDKIAKVKVRLTLREKACQITPHS
 ARSKYFGKREVSIGRAVNHISVWDXDLEDTPTITIMANNEVFCVDTAKBK
 KAARLIVPDLVGRVRCNALLYDTQKLPQAMGASGQPSYPAQVFLFKAWBK
 DPMGYSYDTRCFDSTVTERDTEBESYRACLSPEEAHTAIHSLTERLYVGMFNSK
 GQTCYRRGRASGLVTSTMGNTITCYKALAAKAAHTIATMLVCGDLDLVSBSQ
 TDEEDENLAPTEAMTRVSPAPGPPRDELELITSCSNVSVFSGPQRRYYILTR
 DPTTILARAWEVTRHSVNSMLGNIIOVAPTIVARMVLMTHFSILMAQDITDLN
 FEMVGVSVSPLDLPAILERLGLDAPSLHTYTPHETRVASALRKGAPPLRWKS
 RARVASLISRGGAACGVFLFNWAKTKLTLTPPEARLLDLSWFTVAGGGDI
 YHSVSPARPLLGLLGLFVCGGLFLLPAR

BASE COUNT 1970 a 2820 c 2633 g 2166 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0 Length: 9589
 Score: 15830.00 Matches: 2967
 Percent Similarity: 98.81% Conservative: 30
 Best Local Similarity: 97.82% Mismatches: 36
 Query Match: 98.20% Indels: 0
 DB: 14 Gaps: 0

US-09-980-559-2 (1-3033) x HPCPOLP (1-9589)

Qy 1 MetSerThrAspPolsProGlnAArglyTherlySerGlnThrAsnArgProGln 20
 Db 341 ATGAGCAATCTCTAACTCTAAGAAACCAAGAACACCCACCTCGCCCA 400
 Qy 21 AspVallyPheProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 40
 Db 401 GACGTTAAGTTTCGCGCGCGCGCCAGATCGTTGCGGAGTATACTTGTTC 460
 Qy 41 GlyProArgLeuGlyValArgAlaThrArglySerGlnArgGlyProArgGly 60
 Db 461 GCGCCAGGTGGGTGGCGCGGCAAGCAAGACTTCGAGCGGTCCACGACGTTGA 520
 Qy 61 ArgArgGlnProIleProIlePheArgArgSerThrGlyLysSerThrGlyLysProGly 80

Copied from 09540843 on 05/19/2004

[illegible]

